The Craft of Sociological Research

The Craft of Sociological Research

Principles and Methods of Collecting, Analyzing, and Presenting Social Science Data

VICTOR TAN CHEN; GABRIELA LEÓN-PÉREZ; JULIE HONNOLD; AND VOLKAN AYTAR

VIVA, VIRGINIA'S ACADEMIC LIBRARY CONSORTIUM
RICHMOND, VIRGINIA



The Craft of Sociological Research Copyright © by Victor Tan Chen; Gabriela León-Pérez; Julie Honnold; and Volkan Aytar is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License, except where otherwise noted.

Contents

Using This Textbook 1

Copyright Information 3

Suggested Citation v

Acknowledgements vii

Lists of Sidebars, Figures, Tables, and Videos ix

Sidebars ix

Figures xii

Tables xvii

Videos xviii

1. Introduction

1.1. The World according to Science 25

1.2. Sociology as a Social Science 33

1.3. Design and Goals of This Textbook 41

1.4. References 43

2. Using Sociology in Everyday Life

2.1. Sociologists as Scholars 49

2.2. Sociologists as Market Researchers 55

2.3. Sociologists as Policy Advisors and Evaluators 59

2.4. Sociologists as Clinicians and Public Health Researchers 69

2.5. Sociologists as Activists and Change Agents 73

2.6. Sociologists as Professionals and Citizens 79

2.7. References 83

3. The Role of Theory in Research

3.1. What Do We Mean by “Theory”? 87

3.2. Levels of Analysis 89

3.3. The Components of Theories 97

3.4. Evaluating Theories 119

3.5. Major Sociological Theories and Paradigms 125

3.6. Linking Theory and Data 135

3.7. References 137

4. Research Questions

4.1. Choosing a General Research Topic 141
4.2. Types of Research Questions 157
4.3. Posing Explanatory Research Questions 163
4.4. Refining Your Research Question 173
4.5. References 179

5. Research Design

5.1. What Should My Literature Review Cover? 183
5.2. What Sources Should I Use? 189
5.3. How Do I Find Relevant Sources? 201
5.4. How Do I Read and Evaluate Sources? 209
5.5. How Do I Write a Literature Review? 219
5.6. How Do I Write a Research Proposal? 229
5.7. References 239

6. Sampling

6.1. Units of Analysis 245
6.2. Populations versus Samples 251
6.3. Probability Sampling 263
6.4. Nonprobability Sampling 277
6.5. Questions to Ask about Samples 295
6.6. References 299

7. Measuring the Social World

7.1. Operationalizing a Concept 305
7.2. Deciding on the Correct Level of Measurement 315
7.3. Indexes and Scales: Measures with Multiple Items 325
7.4. Measurement Quality 333
7.5. References 353

8. Ethics

8.1. Subject to Harm: Coercion, Deception, and Other Risks for Participants 359
8.2. The Belmont Principles: Respect for Persons, Beneficence, and Justice 371
8.3. Regulating Research: Institutional Review Boards and Professional Norms 387

8.4. Veracity and Objectivity in Research 393
8.5. References 405

9. Ethnography

9.1. When and How to Use Ethnography 413
9.2. Going into the Field 427
9.3. Field Jottings and Field Notes 443
9.4. References 459

10. In-Depth Interviews

10.1. When Should In-Depth Interviewing Be Used? 467
10.2. Creating Interview Guides 471
10.3. Preparing for In-Depth Interviews 481
10.4. Conducting In-Depth Interviews 487
10.5. Focus Groups 497
10.6. References 503

11. Qualitative Data Analysis

11.1. Coding Qualitative Data 509
11.2. Creating Arguments out of Codes 529
11.3. Writing Up Qualitative Results 539
11.4. References 549

12. Experiments

12.1. Pre-experimental Designs: What You Should Not Do 555
12.2. Randomized Controlled Designs: What You Should Do 565
12.3. Persistent Validity Problems: What You Still Need to Avoid 571
12.4. Field Experiments: External Validity at the Cost of Control 587
12.5. Quasi-experiments: Approximating True Experiments with Observational Data 597
12.6. Experiments: Our Best Way of Inferring Causality, but Far from Foolproof 615
12.7. References 619

13. Surveys

13.1. The Strengths and Weaknesses of Survey Research 625
13.2. Types of Surveys 633
13.3. Designing Effective Questions and Questionnaires 645
13.4. References 659

14. Quantitative Data Analysis

14.1. Preparing for Data Analysis 665
14.2. Univariate Data Analysis: Analyzing One Variable at a Time 683
14.3. Bivariate Data Analysis: Crosstabulations and Chi-Square 707
14.4. References 737

15. Materials-Based Methods

15.1. Content Analysis 743
15.2. References 767

Appendix A: Presenting, Writing, and Publishing

A.1. Deciding What to Share and with Whom to Share It 771
A.2. Presenting Your Research 777
A.3. Writing Up Your Work for Other Scholars 789
A.4. Reaching Your Intended Audience 821
A.5. References 829

Glossary 831
Credits 869
1. *Introduction* 869
2. *Using Sociology in Everyday Life* 870
3. *The Role of Theory in Research* 872
4. *Research Questions* 875
5. *Research Design* 876
6. *Sampling* 877
7. *Measuring the Social world* 879
8. *Ethics* 881
9. *Ethnography* 883
10. *In-Depth Interviewing* 884
11. *Qualitative Data Analysis* 885
12. *Experiments* 887
13. *Surveys* 890
14. *Quantitative Data Analysis* 891
15. *Materials-Based Methods* 893
A. *Presenting, Writing, and Publishing* 894
Contributors 897

Using This Textbook

Welcome to *The Craft of Sociological Research*! We hope you find this openly licensed textbook on research methods informative and easy to use.

Reading This Textbook on a Web Browser

If you are reading this online, you can navigate between chapters and between sections easily by clicking or tapping on the black **Contents** drop-down menu at the top of the page. A list of all the textbook's chapters will appear. Click or tap on the **plus (+)** buttons to the right of each chapter title to see the sections for that chapter, which you can then click or tap to access.

In addition to using the contents menu, you can flip quickly through the textbook by clicking or tapping on the **red bars** that appear at the bottom of your screen. These links allow you to move to the next section of the textbook or return to the previous section. The **red arrow** at the bottom of the screen will take you to the top of the page.

Click or tap on any underlined terms to see their **glossary** definitions. Any links embedded in the names of **chapters**, **sections**, **figures**, and **tables** in the text will take you directly to that material.

Note that you can also install the free Hypothesis extension in your web browser, which will allow you to highlight specific passages of the text, insert notes, and share your annotations with anyone you wish—which may be useful for student study groups, among other things.

All set? Head over to the Introduction to start reading.

Other Ways of Reading This Textbook

In addition to reading the textbook in a web browser, you can download it as a single file that can then be read on any device or printed on paper. Use the links below to download the full text for free.

Type of File	Best For	Notes
PDF for web	Reading on a computer, tablet, or phone using a PDF reader	Includes links and accessibility tags
PDF for print	Printing a hard copy	Does not include links
EPUB	Reading on iBooks, Nook, and other e-reading apps and devices	

Note: Other file formats (e.g., MOBI, Pressbooks XML, Common Cartridge) will be offered shortly.

Supplementary Instructional Materials and Textbook Listserv

We have compiled a number of free instructional materials at the textbook's supplementary website at sites.google.com/view/sociology-research-methods. These include lists of relevant datasets, journals, and professional associations, career advice, a style guide, and guidelines for doing the SPSS statistical analysis described in the textbook.

In the near future, we also plan on providing lecture slides and videos. To stay in the loop about any updates, feel free to subscribe to our low-volume mailing list at inthefray.org/craft. Updates will also be available at victortanchen.com/books/the-craft-of-sociological-research.

Providing Feedback and Contacting the Authors

We are continually revising and expanding on this textbook. Please do not hesitate to let us know about any errors, however major or minor. We are also open to suggestions about topics to cover, examples to include, and other ways to improve the material. Feel free to email the coauthors at craft@inthefray.org or use our textbook's open Hypothesis group to highlight specific passages and make comments directly in the online text. We regularly review this feedback.

Copyright Information

The Craft of Sociological Research: Principles and Methods of Collecting, Analyzing, and Presenting Social Science Data was created by Victor Tan Chen, Gabriela León-Pérez, Julie Honnold, and Volkan Aytar. Unless otherwise noted, this textbook is © 2024 by these authors and licensed under a **Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License** (https://creativecommons.org/licenses/by-nc-sa/4.0).



This means you are free to:

- **Share**–copy and redistribute the material in any medium or format
- **Adapt**–remix, transform, and build upon the material

If you adapt the textbook or create derivative resources, you must follow these license terms:

- **Attribution**–You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your uses.
- **Noncommercial**–You may not use the material for commercial purposes.
- **Share Alike**–If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.
- **No Additional Restrictions**–You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

Suggested Citation

The following is a suggested citation for the textbook in American Sociological Association (ASA) style:

> Chen, Victor Tan, Gabriela León-Pérez, Julie Honnold, and Volkan Aytar. 2024. *The Craft of Sociological Research: Principles and Methods of Collecting, Analyzing, and Presenting Social Science Data.* Fairfax: Virtual Library of Virginia. Available at: https://viva.pressbooks.pub/sociology-research-methods

Suggested Attribution

For content creators who wish to adapt the content in our textbook, please use the following attribution:

> *The Craft of Sociological Research: Principles and Methods of Collecting, Analyzing, and Presenting Social Science Data*, by Victor Tan Chen, Gabriela León-Pérez, Julie Honnold, and Volkan Aytar, distributed under a CC-BY-NC-SA 4.0 license.

This textbook incorporates openly licensed content from other sources, and a list of attributions is provided at the end of the book in the **Credits** section.

Acknowledgements

Our textbook draws on the work of a community of scholars, educators, students, and experts in open education, graphic design, publishing, and beyond. We are immensely grateful to everyone involved. We began working on this project at the start of the Covid-19 pandemic, and we would not have been able to reach this milestone of publication without the collective work and support of these individuals.

First, this textbook would not have been possible without the financial assistance of the Virtual Library of Virginia (VIVA) and VCU Libraries. We are indebted to Jessica Kirschner, VIVA's digital publishing coordinator and formerly the open educational resources (OER) librarian at VCU Libraries. In those roles and as our project's open educational resources, publishing, and copyright consultant, Jessica provided invaluable guidance and thoughtfully shepherded the book from an idea in a grant proposal to a fully realized publication. We would also like to acknowledge Stephanie Westcott, VIVA's open and sustainable learning coordinator, for her unflagging support of the project.

We are grateful to Bizhan Khodabandeh, an assistant professor at VCU's Robertson School of Media and Culture, for his creativity and vision. As the textbook's graphic designer, Bizhan created the visual design for the textbook and all the figures in the chapters. We deeply appreciated his collegiality, professionalism, and dedication throughout this project.

Special thanks to Tom Woodward, director of learning spaces and technologies at Middlebury College. As our technology lead, Tom gave us cheerful and expert help in working through various technical issues. Matt Roberts, web developer/designer at VCU Online and Continuing and Professional Education, also provided key technical assistance with Pressbooks and the online publishing process.

A huge thanks goes to Matt DeCarlo, who was amazingly generous in his advice and guidance as we began this undertaking. His enthusiasm for OER was infectious, and we benefited greatly from his example and from the openly licensed material in his two excellent research methods textbooks: *Scientific Inquiry in Social Work* (2018) and *Graduate Research Methods in Social Work* (2021).

We would also like to thank Maureen O'Driscoll, copy editor, and Suzanne Pfister, project manager, at KnowledgeWorks Global Ltd. for their incredible copyediting support. Andrew Park helped locate and organize citations and references and also wrote a sidebar, ***The Bell Curve* and Its Critics**. Many thanks to Melanie Barnes, senior finance technician at VCU's College of Humanities and Sciences, for her administrative assistance in managing our grants. Thanks as well to Kelsey Cheshire and Abbey Childs at VCU Libraries for their support and suggestions.

We are indebted to faculty and students who reviewed and provided constructive feedback on chapters: Katherine K. Chen, Eli Coston, Meredith Katz, Frankie Mastrangelo, and Ian Almond.

We are incredibly grateful to the sociologists who allowed us to interview them for this textbook and whose expertise provides real-life examples of the methodological concepts and approaches explained in many of our chapters: Elyas Bakhtiari, Amy L. Best, Didem Danış, Sarah Dewees, Jennifer Johnson, Leslie Martin, Chinyere Osuji, Allison Pugh, and Joyce Rothschild.

Finally, we would like to express our great appreciation to the authors whose open textbooks served as content sources for many of our chapters: Anol Bhattacherjee, Amy Blackstone, I-Chant A. Chiang, Matthew DeCarlo, and Rajiv Jhangiani.

Lists of Sidebars, Figures, Tables, and Videos

Sidebars

By chapter and section number:

1. Introduction

1.0. The Language of Science

1.1. Deeper Dive: Ontology and Epistemology

2. Using Sociology in Everyday Life

2.1. Researchers and Research Communities

2.3. Finding Sociological Jobs

2.5. Sociology for the People: A Q&A with Sarah Dewees

3. The Role of Theory in Research

3.2. Sociology across Boundaries: A Q&A with Chinyere Osuji

3.3. Deeper Dive: Revising Our Social Realities

3.3. Deeper Dive: Idiographic and Nomothetic Explanations

4. Research Questions

4.1. Asking Questions in Need of Answers: A Q&A with Joyce Rothschild

5. Research Design

5.2. Peer Review: What Makes a Science a Science

5.3. Search Operators

5.4. A Hierarchy of Evidence?

5.4. How to Read Social Science Writing Strategically

5.6. Mapping the Social World: A Q&A with Jennifer A. Johnson

6. Sampling

6.1. Deeper Dive: Ecological Fallacies

6.2. Deeper Dive: Sampling Gone Wrong

6.4. Deeper Dive: Sampling Gone Wrong (Again)

7. Measuring the Social World

7.1. *The Bell Curve* and Its Critics, by Andrew Park

7.2. Deeper Dive: Interval and Ratio Levels of Measurement

7.4. Social Desirability in Political Polling: The Wilder Effect and Shy Trump Voters

8. Ethics

8.1. The Nuremberg Code

8.2. The Journalist and the Ethnographer, by Victor Tan Chen

8.3. Professional Codes of Ethics

8.4. Studying Vulnerable Populations and Polarized Issues: A Q&A with Didem Danış

9. Ethnography

9.2. Ethnography as Empathetic Understanding: A Q&A with Amy L. Best

10. In-Depth Interviews

10.4. Capturing Stories that Resonate: A Q&A with Allison Pugh

11. Qualitative Data Analysis

11.0. Transcribing Interview Data

11.1. Deeper Dive: Axial and Selective Coding

11.1. Actionable Research: A Q&A with Leslie Martin

12. Experiments

12.3. Deeper Dive: Problems with Pretests

12.5. Deeper Dive: Time-Series Design

13. Surveys

13.1. Response Rates and the Dirty Secret of Political Polling

14. Quantitative Data Analysis

14.0. Statistical Analysis Programs

14.1. Deeper Dive: Entering Your Own Data in SPSS

14.1. Searching a Data Repository

14.2. Deeper Dive: Treating Ordinal-Level Variables as Scale-Level Variables

14.2. Deeper Dive: Measures of Variability

14.2. Quantifying and Comparing Life Outcomes: A Q&A with Elyas Bakhtiari

14.3. General Tips for Reading Data Tables

14.3. Understanding Null Hypothesis Testing

14.3. Interpreting Inferential Statistics

A. Presenting, Writing, and Publishing

A.2. Sociology Professional Conferences

A.2. Deeper Dive: Tips for Presentation Slides

A.3. Deeper Dive: Using Zotero to Manage Your References

A.4. Writing Op-eds, by Victor Tan Chen

Figures

By numbering within chapter:

3.1. Macro, Meso, and Micro Levels of Analysis

3.2. The Theoretical and Empirical Levels of Research

3.3. The Relationships between Concepts

4.1. Example of a Concept Map

4.2. The Wheel of Science

4.3. Deduction: From Abstract to Concrete

4.4. Induction: From Concrete to Abstract

5.1. Google Scholar

5.2. Literature Review Matrix

5.3. Organizing Your Literature Review as a Funnel

6.1. Units of Analysis, Populations and Samples

6.2. Target Populations and Sampling Frames

6.3. Respondents versus Nonrespondents

6.4. Systematic Sampling

6.5. An Example of Periodicity When Using Systematic Sampling

6.6. Examples of Proportionate and Disproportionate Stratified Sampling: Racial and Ethnic Groups in Richmond, Virginia

6.7. An Example of Cluster Sampling: Surveying Public Librarians in Virginia

6.8. Snowball Sampling

6.9. Example of Quota Sampling

7.1. The Conceptualization-Operationalization-Measurement (COM) Process

7.2. Nominal Level of Measurement Example: Dog Breeds

7.3. Ordinal Level of Measurement Example: Perceived Health

7.4. The Hierarchy of Levels of Measurement

7.5. Fahrenheit, Celsius, and Kelvin Temperature Scales

7.6. Social Progress Index

7.7. Comparing Reliability and Validity

9.1. The Participation Continuum

9.2. Organizational Permission Letter

9.3. The Disclosure Continuum

9.4a. Field Jottings

9.4b. Vignette

9.5a. Organizing Field Notes by Event or Organization

9.5b. Organizing Field Notes by Date

9.6. Example of a Field Note

10.1. Interview Guide with General Topics

10.2. Interview Guide with Topics and Questions

10.3. Supplemental Questionnaire

10.4. Asking Sensitive Questions

10.5. Recruitment Flyer

11.1. Automated Transcription Using NVivo

11.2. Coding Using NVivo

11.3. Compiling Coding References

11.4. Interviewee Spreadsheet

11.5. Folder of Coding Memos

11.6. Coding Memo

11.7. Published Manuscript Incorporating Coding

11.8. Nesting Codes

11.9. Two-by-Two Table

11.10. Complex Typologies

11.11. Concept Maps

11.12. Blockquotes in Qualitative Studies

11.13. Introducing Interviewees

11.14. Editing Interview Quotes for Clarity

11.15. Extended Descriptions of Interviewees

12.1. One-Group Posttest-Only Design

12.2. One-Group Pre-Post Design

12.3. Two-Group Pretest–Posttest Design

12.4. A Classic Experiment: Two-Group Randomized Pre-Post Design

12.5. Two-Group Randomized Pre-Post Design with Scores (Math App Example)

12.6. Two-Group Randomized Posttest-Only Design

12.7. Outcomes at 24 Years or Older for Children Who Moved Before They Were 13: An Analysis of Moving to Opportunity Data by Chetty et al. (2016)

12.8. Appointments Obtained for Mental Health Therapy by the Class, Race, and Gender of the Help-Seeker: An Audit Study by Kugelmass (2016)

12.9. The Effect of a Criminal Record on Black and White Job Applicants: An Audit Study by Pager (2003)

12.10. Nonequivalent Control Group Design

12.11. Educational Attainment of Cambodian Students by Dropout Risk Score: Regression Discontinuity Analysis by Filmer and Schady (2009)

12.12. The Logic of the Differences-in-Differences Method in a Minimum-Wage Study by Card and Krueger (1994)

12.13. One-Group Time-Series Design

12.14. Time-Series Outcome When the Treatment Has an Effect

12.15. Oklahoma County General Fertility Rate over Time: A Time-Series Experimental Design (Rodgers et al. 2005)

12.16. Multiple Time-Series Design with Two Comparison Groups

13.1. 1890 Census Form

13.2. Letter with a Survey Link

13.3. Questions for Interviewers

13.4. Question Order Effects

13.5. Mutually Exclusive Response Options

13.6. Collectively Exhaustive Response Options

13.7. Double-Barreled Response Options

13.8. Survey Questions in a Matrix Format

13.9. Filter Questions

14.1. Codebook Excerpt from the LIFE Dataset of the Mexican Migration Project

14.2. Levels of Measurement

14.3. Numerical Codes for a Nominal-Level Variable's Response Options

14.4. Numerical Codes for an Ordinal-Level Variable's Response Options

14.5. Viewing a Dataset in SPSS

14.6. Variables in the GSS Data Explorer

14.7. Variables in the GSS Codebook

14.8. Showing Value Labels for Variables in SPSS

14.9. Listing the Value Labels for a Variable

14.10. Breaking Down a Frequency Table

14.11. Pie Chart for Frequency of Prayer

14.12. Bar Chart for Frequency of Prayer

14.13. Frequency Distribution for a Recoded Age Variable

14.14. Histogram of the Scale-Level Age Variable

14.15. Line Graph of the Scale-Level Age Variable

14.16. Three Frequency Distributions with Their Modes Highlighted

14.17. Modes for Age and Educational Attainment

14.18. Median for Frequency of Prayer

14.19. Histograms Showing Shapes of Frequency Distributions

14.20. Mean for Frequency of Prayer, an Ordinal-Level Variable

14.21. Frequency Table from Lee et al. (2011)

14.22. Crosstabulation of Frequency of Prayer (DV) by Gender (IV) with Counts

14.23. Crosstabulation of Frequency of Prayer by Gender with Column Percentages

14.24. Crosstabulation of Frequency of Prayer by Gender with Row Percentages

14.25. Possible Relationships between Variables in a Crosstab

14.26. Possible Directions of Relationships in a Crosstabulation

14.27. Chi-Square Test for the Crosstabulation of Frequency of Prayer by Gender

14.28. Chi-Square Test for the Crosstabulation of Frequency of Prayer by Educational Attainment

14.29. Table in Kimunai et al. (2016) Presenting Crosstabulations of Healthcare Behaviors by Level of Literacy

15.1. Expert Quotes about Police Shootings in the New York Times and USA Today: A Content Analysis by Crichlow and Fulcher (2017)

15.2. Coding Instructions from Neuendorf et al. (2010)

15.3. Calculating Inter-coder Reliability in Neuendorf et al. (2010)

15.4. Frequencies of Codes in Student Comments: A Content Analysis by Chambers and Chiang (2012)

15.5. Views about Marriage Norms in U.S. Women's Magazines from 1900 to 1979: A Quantitative Content Analysis by Cancian and Gordon (1988)

15.6. The Relationship between Gender and Scouting Activity Type: A Content Analysis by Denny (2011)

15.7. The Relationship between the Gender Category and Color Palette of Toys Sold on the Disney Store Website: A Content Analysis by Auster and Mansbach (2012)

15.8. Challenges in Coding: Pages from Anne Lister's Diaries

A.1. Inserting Formal Hypotheses into Your Paper

Tables

By numbering within chapter:

3.1. Major Sociological Theories and the Study of Sports

3.2. Social Scientific Paradigms

4.1. Predicting the Effects of Arrest on Future Spouse Battery: Testing Theories through Deduction (Sherman and Berk 1984)

4.2. Predicting the Effects of Arrest on Future Spouse Battery: A New Theory Generated from Induction

4.3. Words to Watch for in Research Questions

6.1. When to Use Probability Sampling Strategies

6.2. When to Use Nonprobability Sampling Strategies

7.1. Sociological Examples of a Nominal Level of Measurement

7.2. Sociological Examples of an Ordinal Level of Measurement

7.3. Sociological Examples of a Scale Level of Measurement

7.4. Acceptance of Social Groups Scale

7.5. Confidence in the U.S. Federal Government: Summed Scores for Respondents' Perceptions of the Three Branches of Government

7.6. Confidence in the U.S. Federal Government: Mean Scores for Respondents' Perceptions of the Three Branches of Government

9.1. Examples of Ethnographies

9.2. Telling versus Showing

11.1. Typology of Parenting Approaches in Lareau (2011)

12.1. Threats to Internal Validity and Ways to Address Them

14.1. Measures of Central Tendency for Variables at the Scale, Ordinal, and Nominal Level

14.2. Measures of Variability for Variables at the Scale, Ordinal, and Nominal Level

14.3. Range, IQR, and Standard Deviation for Selected GSS Variables

14.4. Frequency Distribution and Index of Qualitative Variation for Gender and Race

14.5. Crosstabulations of Frequency of Prayer by "Comfort Theory" Independent Variables

14.6. Crosstabulations of Religious Service Attendance by "Comfort Theory" Independent Variables

14.7. The Statistical Significance of a Bivariate Relationship as Predicted by Relationship Strength and Sample Size

15.1. Examples of Content Analysis

15.2. Sample Code Sheet for Study of Kitchen Utensil Popularity over Time

A.1. In-Text Citations

A.2. References in the Works Cited Section

Videos

By numbering within chapter:

3.1. Macro and Micro Perspectives

6.1. How a Flawed Sample Sank a Major Presidential Poll

6.2. Random Sampling

6.3. Probability Sampling Strategies

6.4. Opt-In Surveys

8.1. An Inauspicious Start to an Experiment

9.1. *Sidewalk*, the Documentary

9.2. Street Corner Society, in Retrospect

11.1. A Typology of Parenting Approaches

12.1. Experimenter Expectancy Effects

13.1. The *Up* Documentary Series

13.2. Writing Survey Questions

14.5. Measures of Central Tendency

15.1. What Kitchen Utensils Say about Society

1. INTRODUCTION



Science is a particular way of knowing that systematically investigates, categorizes, and explains facts or truths. A key word here is *systematically*. Conducting science is a deliberate process. Scientists gather information in a way that is organized and intentional and usually follows a set of predetermined steps. They follow these procedures to arrive at an understanding of the world that accurately reflects reality and holds up in the face of scrutiny and criticism.

Of course, you don't have to be a scientist to know how the world works. You don't even need to be a scientist to conduct research. For example, people routinely do "research" on the best place to buy things they want. Journalists conduct "research" on topics of public interest, such as interviewing people about their views on politics. Students do online "research" to find the information they need to complete their term papers. Businesses conduct "research" on their customers to learn about their spending habits. None of this research, however, can be considered *scientific* unless it (1) contributes to a larger body of scientific knowledge, and (2) follows the scientific method. **Scientific knowledge** refers to a systematic and organized body of knowledge that is acquired using the scientific method. The **scientific method** refers to a standardized set of techniques used to acquire and vet knowledge, such as procedures for making observations or interpreting results.



A crowd watches traditional dancers in the Philippines. *Mio Advincula, via Pexels*

Sociology is a *social science*. By "social," we mean that sociology focuses on how human beings interact with one another. By "science," we mean the use of procedures—as we've noted, organized and intentional procedures—to uncover truths about that behavior. Specifically, **sociology** *is the study of social life and social groups*. Sociologists are interested in how individuals shape—and are shaped by—the social groups around them, which may be as small as individual families or couples, or as large as whole nations. Put another way, sociologists help us understand how our personal lives are influenced by larger social forces, such as social networks, institutions, and cultures—in short, who we know, what rules guide our actions, and how the various groups we are part of view the world.

Compared with other social sciences, sociology is a broad field that uses a variety of strategies and techniques to gather and assess scientific knowledge. We call these systematic approaches to learning about our social world **research methods**. This textbook is devoted to teaching you about sociological research methods. It is geared to students at all levels, with "Deeper Dive" sections that are intended for more advanced students as well as sidebars that can also be skipped by readers who just want the basics (the first of those sidebars, **The Language of Science**, follows with a breakdown of key terms). After reading this book, we hope you will be better able to understand the results of research studies, critique their approaches and conclusions, and conduct rigorous research of your own.

In this chapter, we'll consider the variety of ways that science allows us to know things. We'll also talk about what makes *social scientific* research—and *sociological* research—unique in relation to other forms of understanding reality.

The Language of Science

All scientists employ jargon—special words that are used by a profession or group—in the course of their work. By reading and discussing each other's work, scientists ensure the quality and relevance of their research. Using a shared language, in turn, helps create a strong scientific community that can draw on a common identity and common set of perspectives—the basis for a constructive dialogue.

It's important for students new to a scientific field to know its terms and how they are appropriately used. A critical and occasionally prickly bunch by temperament, scientists may dismiss outright the views of people who are unfamiliar with their lingo and traditions. To help you familiarize yourself with the language of science, we've listed some key terms used virtually everywhere across scholarly and scientific fields. Note that not everyone agrees on the meaning of these terms—or any of the terms used in this book, frankly—so perfectly respectable scientists may disagree with what we say here (but hopefully not.)

Scholars, researchers, scientists. "Scholar" can be used as a synonym for words like "researcher" and "scientist," although it typically refers to a person who engages in research for its own sake (rather than for a practical purpose, as we'll discuss later). A college professor would be a scholar. Nonscientists who are academics (such as a literature professor) would also be scholars.

Disciplines, fields, subfields. Scholars do their work within an academic "field" or "discipline." Sociology is an academic field. Each discipline has its own traditions, philosophies, and research methods. Sometimes the distinction seems arbitrary—for instance, "history" is a particular research method (the historical method), whereas "political science" focuses on a specific domain of human activity. Additionally, there is plenty of overlap between fields: sociologists study politics, anthropologists study social ties, economists study culture, and so on. That said, because advances in knowledge within a field often happen through interactions between scholars within a particular field, it's important to understand and appreciate these disciplinary boundaries. Each field also has subfields—for example, economic sociology and social psychology are two subfields within sociology.

Phenomenon. Scientists are typically interested in studying **phenomena**, a catchall term for things, events, and processes that (1) can be observed and (2) raise interesting questions for scientists about how the world works. Sociologists and other social scientists study *social* phenomena, which have to do with human beings and their interactions.

Observation. Scientists don't just sit in an ivory tower and ruminate about the world (or at least ideally, they don't). They draw upon actual experiences seeing, hearing, or otherwise documenting the reality around them, which should be the basis for any knowledge they produce. These observations can be the researcher's own or those of other people, but in either case, the observations must be conducted in a systematic and credible way.

In addition to this general meaning, the word "observation" has more specific uses in the social sciences. As we will discuss in **Chapter 9: Ethnography**, it can refer to **ethnographic observation**—a researcher's direct observation of social spaces as a method of gathering qualitative data. As we'll talk about in **Chapter 12: Experiments**, scientists also distinguish between **experimental data**—information collected under experimental conditions, where researchers can control the conditions under study—and **observational data**,

which scientists gather in the real world from surveys, in-depth interviews, ethnographic observation, and other methods that don't use an experimental design.

Data, datum, data point, dataset. We'll use the word **data** quite a bit in this textbook. It refers to facts, details, statistics, and other information used to understand some phenomenon of interest. Ideally, data is collected through a scientific process of the sort described in this chapter, but any information could be considered data—what matters for the strength of a particular study is often *how* the scientist uses the data. You'll notice that we have referred to "data" as a singular noun. This is a matter of great controversy among scientists. (Did we mention scientists are a bit judgmental?) Regularly, they refer to "data" in the plural form ("The data are quite conclusive"), even though no normal people ever do this. They point out, pedantically, that "datum" is the singular form of data. At the risk of alienating our colleagues, we will refer to "data" in the singular form. In doing so, we follow the lead of many mainstream news publications (Rogers 2012), and since one theme of this book is the importance of connecting research to the real world, we humbly suggest this simple way of making your writing more accessible to nonscientists. That said, when talking to other social scientists, you may find it helpful to **code-switch** and show you are a card-carrying scientist by using "data" as a plural noun (we forgive you).

In the scientific community, researchers use the term **data point** to refer to a specific statistic or other piece of information ("That data point is an outlier"). Now let's consider the opposite case: lots and lots of data. All the data relating to a specific survey or study can be grouped together into a database, or what scientists call a **dataset**. For example, think about a list of students in your class, which includes information about their names, ID numbers, and email addresses. That would be a (small) dataset, but social scientists regularly work with datasets that contain information about thousands of people, with specific details about each individual's personal background and attitudes on a variety of subjects. Note also that the data in datasets doesn't have to deal with individuals per se: as we will describe later, it can be about anything—from books and movies, to families and relationships, to organizations and countries.



Even the notoriously antisocial physicist Isaac Newton (1642–1727) recognized his intellectual debts to the scientists who preceded him. *Wikimedia Commons*

Literature. Science always builds on existing knowledge. Even if their scientific research overturns parts of that accumulated knowledge, scientists should always know what previous studies have been done on their subject area, and they need to connect their own work to that body of knowledge. **Literature** is the name we give to the existing scientific knowledge relating to a particular phenomenon. Here, we're not talking about the "literature" you read in your English classes—novels and poetry and the like. The *scientific* literature amounts to all the articles, books, and other material written by scholars on a specific topic. Scientists (and scholars more broadly) care so much about literature that they bog down their writing with endless parenthetical citations (a topic discussed in **Appendix A: Presenting, Writing, and Publishing**). Their intention is not to bore their readers, but rather to precisely source their information and be clear about whatever intellectual debts they owe to other scholars.

Popular culture often depicts scientists as antisocial loners and mad geniuses (which, to be honest, is not always an unfair assessment). But the fact that they are deeply rooted in the scientific literature on their subject ensures that they are contributors to a grand tradition of learning, rather than just unwashed crackpots toiling away in their garage laboratories. "If I have seen further it is by standing on the shoulders of Giants," the physicist Isaac Newton famously said (Turnbull 1959:416). Reading and connecting your work to the existing literature on your topic is how you, as a scientist, stand on the shoulders of the scientists who came before you.

Opening chapter image credit: *Danilo Ugaddan, via Pexels. Adapted by Bizhan Khodabandeh.*

1.1. The World according to Science

Learning Objectives

1. Identify and describe the various ways of knowing.
2. Understand the weaknesses of nonsystematic ways of knowing.
3. Understand the role of theories in science and their relationship to data.

Think about what the conventional family with kids looked like in the middle of the last century. Perhaps what comes to mind is a mother staying at home to raise the kids (a "homemaker") and a father working outside the home (a "breadwinner")—in other words, a household much like the middle-class family of June and Ward Cleaver in the popular 1950s TV show *Leave It to Beaver*. (For you younger readers who don't get that reference, think Marge and Homer Simpson in the animated series *The Simpsons*—a not-so-subtle parody of the Cleaver household.) Of course, families look very different today, now that women are in the paid labor force in numbers comparable to men, and parents of all genders are raising kids. But the "traditional" family was a nuclear family with a male breadwinner. Right?





The traditional nuclear family as depicted in *Leave It to Beaver* (1957–1963) and *The Simpsons* (1989–forever). *Universal / Walt Disney Company*

Well, yes, and no. As historians and sociologists have pointed out, the reality of the time was more complex.[1] The percentage of white children born in 1950 who were actually raised throughout their childhood within *Leave It to Beaver/Simpsons*-style households—by married parents, with only the husband working outside the home—was

1. An influential early book in this line of scholarship was historian Stephanie Coontz's provocatively titled book, *The Way We Never Were* (1992).

just 26 percent, according to an analysis of U.S. census data by Donald Hernandez (1993), a scholar working in the sociological subfield of demography. (For African American children, the percentage was less than 10 percent, Hernandez estimates.)That doesn't mean that families like the Cleavers were rare. About half of white kids and a fifth of African American kids were born into such families in 1950. But unemployment, divorce, and the flitting of some "homemakers" in and out of the paid workforce meant that the "traditional" model was always more of an ideal than a day-to-day reality—even in its heyday. As the sociologist Andrew Cherlin (2014:96) notes: "Beyond doubt, these estimates show that it's an exaggeration to assert that most children lived in breadwinner-homemaker families in the 1950s. But they also suggest that it's inaccurate to conclude, as some writers have, that the breadwinner-homemaker family was a myth."

Among other things, sociologists study widespread assumptions that people take for granted, like the fact that most families in the 1950s were just like the Cleavers. Sometimes, they find that these "common-sense" views—also known as *folk beliefs*—are correct. Other times, they find the opposite. As we've just seen, in spite of all the nostalgia that many people still have for the way families "used to be," what family life was really like decades ago was more complex. What everyone seems to know to be true wasn't so true after all.



Flick freezes his tongue to a flagpole as Ralphie (informally) observes in *A Christmas Story* (1983). *Time Warner*

Of course, you can know things without having a background in science. To make sense of the world, we draw upon our experiences seeing, hearing, or otherwise recording what happens around us. In addition to our firsthand experiences, or **observations**, we can rely on other people's observations. For example, if you grew up somewhere cold, you may have learned that your tongue will stick to a flagpole if it's very cold outside. If you were fortunate, your friend made this observation before you, allowing you to make use of their knowledge without having to spend a substantial amount of time attached to a metal pole while waiting for emergency personnel to extract you.

In both of these scenarios, however, the observation process wasn't really deliberate or formal. Instead, you came to know what you believe to be true through **informal observation**. The problem with informal observation is that sometimes it is right, and sometimes it is wrong. Without any systematic process for observing or assessing the accuracy of our observations, we can never *really* trust our informal observations. This is even more the case when we're dealing with aspects of our reality that are social, rather than physical. Social life is exceedingly complex because human beings are so diverse and so hard to predict or understand.

For example, suppose a friend of yours declared that "all men lie all the time" shortly after she'd learned that her boyfriend had cheated and made up a cover story. The fact that one man lied to her in that instance came to represent all experiences with all men. But do *all* men really lie *all* the time? Of course not. If you prompted your friend to think about her experiences with men more broadly, she would likely acknowledge that she knew many men who, to her knowledge, had never lied to her, and that even her boyfriend didn't generally have a habit of lying.

This friend committed an error of logic that is called **selective observation**. She cherry-picked the data she used to prove her point, noticing only those bits of information that confirmed the pattern she wanted to find. Even if she *didn't* selectively assemble her data, however, we would have good reason to be skeptical about her adamant claims that all men are liars. Given that she is drawing from just one person's observations—her own—your friend is likely to be committing an error of **overgeneralization**, too: assuming that broad patterns exist based on limited observations. Even

if all the men in her life happened to hail from some wretched hive of scum and villainy, we could easily imagine how other people, interacting with other men, might have different luck. The men she knows aren't necessarily good stand-ins for the entirety of that particular gender.

Given the casual, unscientific way that your friend gathered her evidence, the data she generated from her selective and limited observations amounts to **anecdotal evidence**. It is based on stories (anecdotes) from personal experiences. It has not been collected systematically in a way that makes us confident that what was observed reflects what actually happens in most situations. This is concisely stated in the adage: "The plural of anecdote is not data."

Another way that people justify what they believe is by referring to **tradition**: *This is what we've always known to be true*. There's an oft-told tale about a woman who would cut off both ends of a ham before cooking it (Mikkelson 2005). That's the way her mother did it, so clearly that was the way it was *supposed* to be done. Her own child started doing the same thing until they thought to ask the woman's mother why. The only reason, her mother explained, was that her roaster had been too small to cook the ham in one piece.

Without questioning what we think we know, we may wind up believing things that are actually false. This is most likely to occur when an **authority** tells us that something is so. It doesn't matter whether that authority is a government official, a business or religious leader, a teacher or professor, or a ham-hacking mom—just because someone says something is true doesn't mean we need to take their word for it.



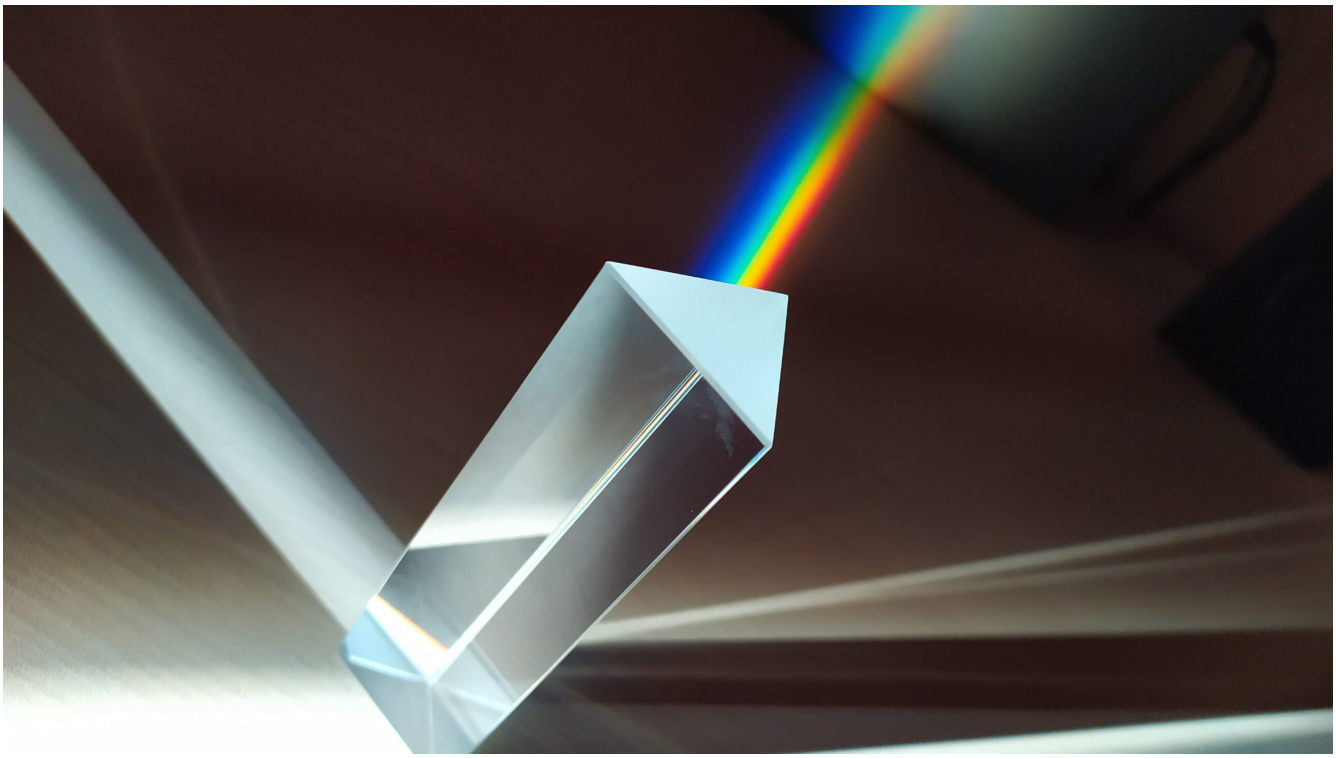
What if everything you knew about roasting hams was wrong? *Tim Sackton, via Flickr*

While it's understandable that people will believe something because someone they look up to said it is so, this way of knowing differs sharply from a *scientific* way of knowing. Beliefs are not necessarily vetted—confirmed or disproven—by independent researchers. They may have originally arisen from selective observations or overgeneralizations. The authorities who are putting forward these beliefs might have committed errors in logic. Sometimes, sadly, they may have an interest in misleading people.

To avoid putting forward such a flawed picture of reality, sociology relies on a systematic process for making and interpreting observations and thereby building scientific knowledge. We'll discuss this process in more detail in the next few sections.

Although sociologists typically have to rely on evidence that goes beyond what people merely believe to be true, beliefs can figure into their work in other ways. Sociologists may examine beliefs to understand what they are and where they come from, for example. However, beliefs in these cases are phenomena *to be studied*—not the evidence that sociologists use to support their conclusions. For example, when they were interviewing low-income fathers in the late 2000s, the sociologists Kathryn Edin and Timothy Nelson (2013) found that many of the dads they talked to still mentioned the parents in *Leave It to Beaver*—a television show that last aired original episodes in 1963—as role models for what they envisioned to be an ideal family. As we described earlier, families like the Cleavers were less common in that earlier period than many people today believe. But the reasons that individuals hold onto such romanticized views of the past is itself worth studying and understanding.

The Scientific Method: Theory and Observation



Much like the theory of optics explains phenomena related to light, sociological theory explains social phenomena—how individuals and groups behave in societies. *Dobromir Hristov, via Pexels*

In its simplest terms, scientific knowledge amounts to a collection of **theories:** systematic explanations of specific phenomena. (We'll have a lot more to say about theories in **Chapter 3: The Role of Theory in Research**.) In physics, for example, scientists have developed theories to explain *physical* phenomena—the nature and workings of the physical reality we can observe. The theory of optics explains the properties of light and how it behaves in different media. Electromagnetic theory explains the properties of electricity and how to generate it. Quantum mechanics explains the properties of subatomic particles. Thermodynamics explains the properties of energy and mechanical work

In sociology, our theories explain social phenomena—people's thoughts and behaviors as individuals and groups acting within societies. For instance, *reference group* theory describes how people judge themselves according to the social groups they relate to or aspire to join (Merton 1957). The theory of *emotional labor* captures how employees are trained to behave a certain way to convey their professionalism and improve their on-the-job performance (Hochschild 2012). *Stress and coping theory* emphasizes how negative experiences (such as discrimination) and social and personal coping resources (such as a person's resiliency, or the social support from friends and family they can draw on) interact and shape people's health outcomes (Lazarus and Folkman 1984).

Sometimes a distinction is made between laws and other theories. **Laws** are regularly observed patterns of phenomena. For instance, in physics, Newton's laws of motion describe what happens when an object is in a state of rest or motion (Newton's first law), what force is needed to move a stationary object or to stop a moving object (Newton's second law), and what happens when two objects collide (Newton's third law). Collectively, the three laws constitute the basis of

classical mechanics—a theory of moving objects. In sociology, we don't typically talk about "laws" because human beings don't generally display such deterministic tendencies: we tend to speak instead about what individuals or groups are *likely* to do, rather than what they *will* do.

Regardless of whether we're talking about the social sciences or the natural sciences, it is important to note that the way scientists understand "theory" is different from the way that people frequently use the term in casual conversation. When you say you have a "theory" that your professor always looks disheveled for your morning class because they attend wild sociology faculty parties, more than likely, your explanation is based on a hunch rather than actual facts. But a scientific theory may or may not be supported by the available facts; in discussions among scientists, the word "theory" says nothing about the evidence that underlies it. Many scientific theories—such as the theory of evolution and the theory of climate change—have been confirmed by an enormous body of past research using many different types of evidence.

The ultimate goal of scientific research is to discover theories that can explain natural or social phenomena. It is important to understand that this knowledge may be imperfect or even quite far from the truth. There may be good or poor explanations, depending on the extent to which those explanations fit well with reality. The progress of science is marked by our progression over time from poorer theories to better theories—through making better observations, using more accurate measures, and employing more informed logical reasoning.

In the simplest terms, we arrive at scientific theories through a process of obtaining evidence and applying logic. Our evidence comes from careful and methodical observations—ours, or those of other reputable individuals. Scientists call this **empirical evidence**, or data. For our data to be "good," "empirically sound" data, it needs to have been collected using the scientific method (more about this later). After collecting our data, we make sense of it based on what we already know from other theories and other data, making a logical case for our particular theory, which should refine, expand upon, or overturn previous theories.

This theory- and data-driven approach is called the *scientific method*. The scientific method allows researchers to independently and impartially test preexisting theories and prior findings and subject them to open debate, modifications, and enhancements. Research based on the scientific method is supposed to be:

1. **Logical:** We make inferences (draw conclusions) based on logical principles of reasoning.
2. **Confirmable:** The inferences we make should match what was actually observed.
3. **Replicable:** Other scientists should be able to independently repeat the scientific study we conducted and obtain similar, if not identical, results. (This pillar of the scientific method is somewhat contested when it comes to certain types of research, as we'll describe in later chapters.)
4. **Scrutinizable:** The procedures used and the inferences derived must withstand critical scrutiny (**peer review**) by other scientists.

Scientists believe "good science" involves both theory and observation and a continual movement—and balance—between them. They distinguish their work from other kinds of inquiry by emphasizing the mix of creative theory and rigorous research methods they employ. For instance, social scientists might differentiate themselves from journalists by talking about how they build on an existing body of scientific theory and evidence, yet they might also draw contrasts between social science and philosophy, pointing out how their work is rooted in empirical evidence. Here, we should note that good research is good research, and you don't have to have the title of a "scientist" (much less "professor") to do it. Furthermore, social scientists often rely on journalistic and scholarly writing outside the social sciences, borrowing from that work to extend the data or theories they draw upon.

The scientific method is the basis for the wide variety of research approaches used in sociology. These research methods range from laboratory experiments, to surveys, to direct observation. Much of this book is devoted to learning about

these different methods. Before we discuss sociological research methods, however, we should better understand where sociology stands in relation to other sciences.

Deeper Dive: Ontology and Epistemology

Thinking about what you know and how you know what you know involves questions of ontology and epistemology—concepts that philosophers typically talk about, but that are relevant to the work of sociologists as well. Whenever we sociologists begin an inquiry into our social world, we start with some understanding of "what *is*," what can be known about what is, and what might be the best way to learn about what is.

Ontology deals with the first part of these sorts of questions. It refers to one's perspective on the nature of reality. In sociology, a researcher's "ontological position" might shape the sorts of research questions they ask and how those questions are posed. Some sociologists take the position that what is real is in the eye of the beholder and that our job is to understand what or how people think about the world around them. Other sociologists feel that although people can differ in their perception of reality, there is only one *true* reality. (We will have more to say about these different perspectives when we talk about scientific paradigms in **Chapter 3: The Role of Theory in Research**).

Like ontology, **epistemology** has to do with knowledge. But rather than dealing with questions about "what is," epistemology deals with questions of *how* we know what is. In sociology, there are a number of ways to uncover knowledge. We might interview people to understand their opinions on racial discrimination, or we might observe them interact with people of other races and see firsthand how they treat other people. We could avoid face-to-face interaction altogether by sending people surveys to complete independently. We could read what people have already thought and written about this topic by scouring news articles or comments. All these are ways that sociologists gain knowledge. Each method of data collection comes with its own set of epistemological assumptions about how to find things out. We'll talk in more depth about these ways of knowing in our chapters on data collection.

Key Takeaways

1. There are several different ways that we know what we know, including our informal observations, our belief in what tradition or authorities tell us, and our use of scientific research methods.
2. Research methods are a much more reliable source of knowledge than most of our other ways of knowing.
3. Science is a collection of theories that help us explain the world, and these theories are derived from the data we gather—the observations we make about that world.

Exercise

Think about a time in the past when you made a bad decision (e.g., wore the wrong shoes for hiking, dated the wrong person, chose not to study for an exam, paid actual money for a textbook). What caused you to make this decision? How did any of the ways of knowing described previously (**informal observation**, **selective observation**, **overgeneralization**, **anecdotal evidence**, **tradition**, and **authority**) contribute to your error-prone decision-making process? How might you use social scientific research methods to help improve your decision-making in this area in the future? Describe how you might gather empirical evidence and use logic to inform your decision. For example, if you wanted to convince a professor to assign different course materials, you could collect data on the pricing and quality of the various commercial and open-access textbook options for the class. You could use the theory of *cost–benefit analysis* (weighing the strengths and weaknesses of each option against those of alternatives) to make a case for the best option: clearly, convincing your professor that the time spent adapting a course for open-access materials is worthwhile, given the substantial money students would save and the unfathomable joy your professor would feel in contributing to a world of free knowledge for all.

1.2. Sociology as a Social Science

Learning Objectives

1. Understand how sociology is a science and how it differs from other sciences.
2. Understand that social life can be understood as a series of patterns, even if exceptions exist to those patterns.
3. Learn the difference between qualitative and quantitative methods.

Science can be grouped into two broad categories: the natural sciences and the social sciences. The **natural sciences** (sometimes called the "hard sciences") are concerned with naturally occurring objects or phenomena, such as light, objects, matter, earth, stars, or the human body. They include disciplines such as physics (the science of physical objects), chemistry (the science of matter), astronomy (the science of stars and other celestial objects), geology (the science of the earth), and biology (the science of organisms).

In contrast, the **social sciences** are focused on studying people or collections of people, such as groups, firms, societies, or economies, and their individual or collective behaviors. Social sciences can be classified into disciplines such as economics (the science of firms, markets, and economies), psychology (the science of human behaviors), political science (the science of political systems, behaviors, and thought), and sociology (the science of social groups and social life).



In the southern Asian country of Bhutan, the government has made happiness a benchmark for the nation's progress, devising a "Gross National Happiness" survey to track the country's levels of happiness. According to international data, however, Bhutan is still nowhere near the top of the happiness leaderboard—which Nordic countries routinely dominate. *Minh N, via Pexels*

The natural sciences are different from the social sciences in several respects. To a greater extent, the natural sciences have the ability to be precise, deterministic, and independent of the person making the scientific observations. For instance, a scientific experiment in physics, such as measuring the speed of sound or the refractive index of water, should always yield the exact same results, regardless of the time or setting of the experiment or the person conducting it. If two researchers conducting the same physics experiment obtain two different values for these physical properties, then it generally means that one or both of those researchers must be in error. However, the same cannot be said for the social sciences, whose results tend to be more ambiguous and complex.

For instance, if you ask a person how happy they are in a survey, you may find that the same person is more happy or less happy on different days—sometimes, at different times on the same day. Their happiness may vary depending on events that transpired earlier that day. Furthermore, there is not a single method of measuring a person's happiness. (In fact, there are numerous measures of more short-term and more long-term kinds of happiness, which researchers in general call *subjective well-being*). As a result, one set of survey questions may find a person to be "happy," whereas a second set may conclude the opposite. This sort of ambiguity underscores the high degree of measurement error that exists in the social sciences (a topic we discuss in depth later in the book).

Furthermore, there is considerable uncertainty—and little consensus among social scientists—about how research should inform our decisions regarding what laws to pass or rules to establish. For instance, you will not find many disagreements among natural scientists on the speed of light or the speed of the earth around the sun, but you will find numerous disagreements among social scientists on how to address a basic social problem such as poverty. Any person studying the social sciences must be comfortable with the higher levels of ambiguity, uncertainty, and error that regularly confront social scientists, which reflect the high variability and complexity to be found among human beings and across human societies. As the astrophysicist Neil deGrasse Tyson (2016) memorably tweeted, "In science, when human behavior enters the equation, things go nonlinear. That's why Physics is easy and Sociology is hard."

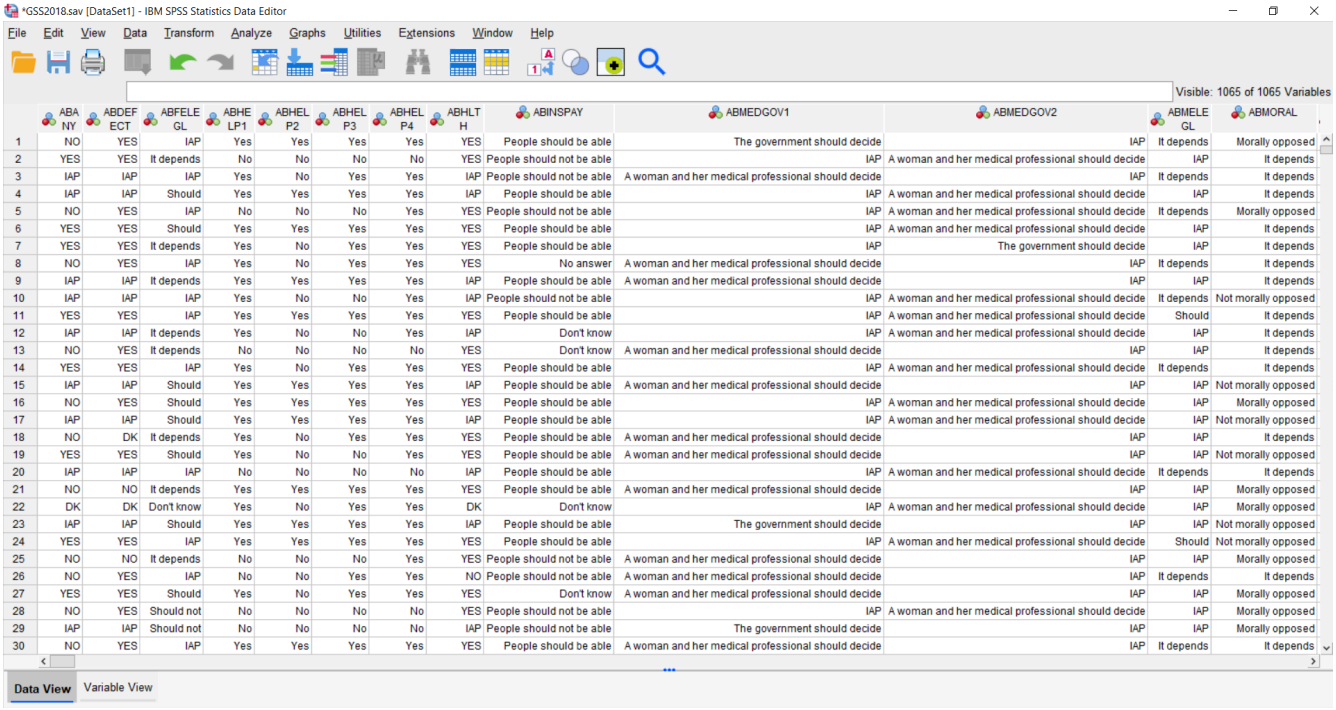


Astrophysicist Neil deGrasse Tyson on why sociology (to paraphrase) is the one science to rule them all. X

Before you get cocky about your powers as a sociologist, remember that the social sciences are also criticized because the inevitable complexity of their findings leaves us with few certain takeaways or readily applicable lessons. A *New Yorker* cartoon once portrayed a little boy looking up at his father while the father tells him, "I'm a social scientist, Michael. That means I can't explain electricity or anything like that, but if you ever want to know about people I'm your man" (Handelsman 1986). This cartoon aptly conveys the distinction between natural and social sciences while slyly poking fun at social scientists and their "soft" science. While we wouldn't say that social scientists are useless because they can't explain electricity, it is true that the conclusions drawn from their research are often so tentative, complicated, and contradictory that they are hard to easily apply to the design of public policies, much less to everyday life.

How Does Sociology Compare to Other Social Sciences?

As we have noted, sociology is a type of social science that focuses on the study of humans in groups. One way that sociology distinguishes itself from other social sciences is its widespread use of both quantitative and qualitative methods. **Quantitative methods** result in data that can be represented by and condensed into numbers. Survey research is probably the most common quantitative method in sociology, but methods such as content analysis can also be conducted in a way that yields quantitative data. **Qualitative methods** are ways of collecting data that yield nonnumerical results, such as words or pictures. Some of the most common qualitative methods in sociology include direct observation, in-depth interviews, and focus groups.



Quantitative datasets like the General Social Survey (pictured here in a screenshot from the SPSS statistical software package) tally up responses from thousands of respondents to give us a representative picture of how the public thinks on any given topic. Every row in the dataset contains the responses of one individual to the various questions posed in the survey.

Both quantitative and qualitative methods study samples of larger populations. As social scientists, we don't usually have data on the entire population we're interested in—say, all Americans or all women—so we collect data from a **sample**, or subset, of that population. The hope is that the members of that sample—what we call **cases**—can appropriately stand in for the whole that we really care about—what we call the **target population**, or **population of interest**. For example, if your target population is students who take sociology methods courses, you probably wouldn't want to just recruit people with high grades for your sample. (We'll have a lot more to say about samples, cases, and populations in **Chapter 6: Sampling**.) Generally speaking, qualitative researchers seek a detailed understanding of a relatively small number of cases. Quantitative researchers are focused less on depth and more on breadth—they typically focus on a much larger number of cases, and they place prime importance on finding samples that adequately represent the target population, which allows them to say with confidence that the results from their sample apply to that larger population as well.

on the road.

Interviewer: Was it a big rig, basically?

Interviewee: Yeah, I wasn't giving anything to the family, really, like I'd pay bills but, really, why do I have a family if I'm never home? And then, like I said, there were some scary nights on the highway, like every day you saw, easy, one if not tons of trucks smashed, especially in the winter. Like you keep driving, they there's times you're going sideways, you're doing whatever, and I'm thinking – it's a crazy industry and after seeing it I can just believe that.

Like in our orientation for trucking, they said the one day right out in front of where we were doing our orientation, 'cause it was a trucking company, this guy going by the place, one that worked for them, he crashed. And they go over there or whatever and the guy was cooking a hotplate, he was cooking his breakfast while he's driving down the road.

Interviewer: Are you serious?

Interviewee: And honestly, there was times when you're sitting there, because all you do is once you get, like once I got the metro airport, cruise controlled it.

Interviewer: What are you picking up? What kinds of loads were you carrying?

Interviewee: They were auto parts, ___ ___ _____, which wasn't good either because now I think he's either out of business. Because he was hooked to Chryslers and it got it a little antsy there.

Interviewer: How many months were you doing that for?

Interviewee: I think I got eight months out of it.

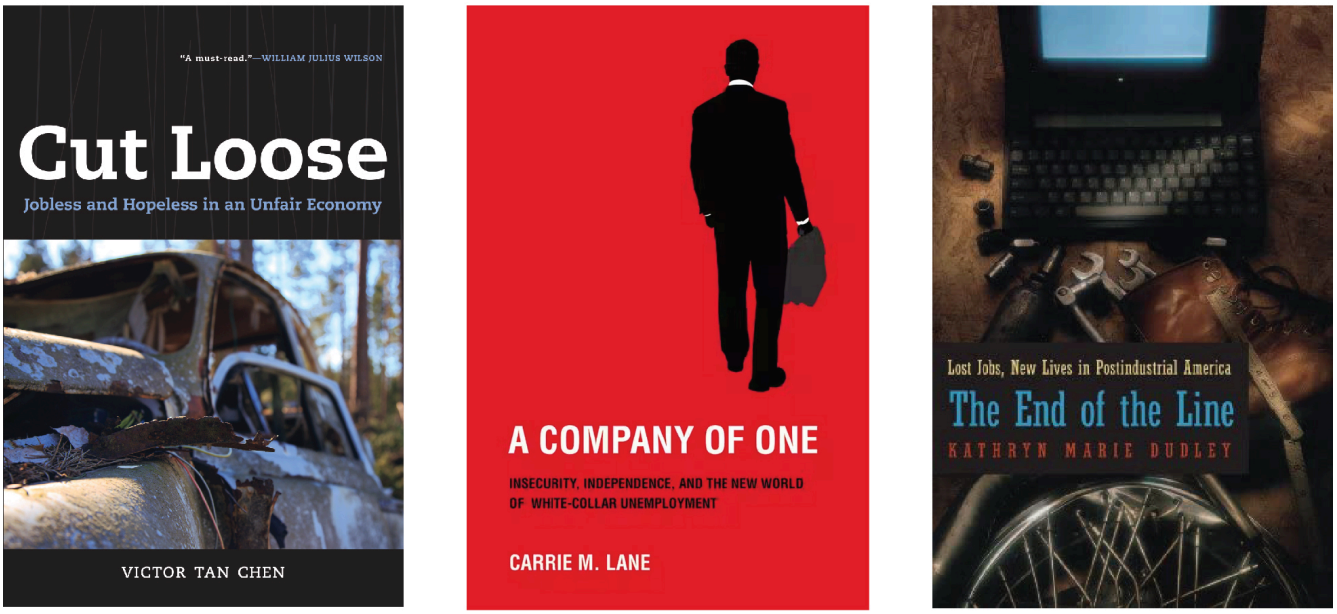
Interviewer: Okay.

Interviewee: Like I probably would have been – I tried to get a local, but you need three years. Most insurance companies don't want you, like if you drive for a bigger company, they can absorb the insurance costs. But they want to make money off you and that means you're going long, long haul.

Interviewer: How many days at a time would you go?

For his book *Cut Loose* (2015), Victor conducted this qualitative interview with an unemployed Canadian worker, who describes here his failed effort to break into the truck-driving industry. (Blank spaces represent words missed in the transcription.)

Sociology has employed quantitative methods from its earliest days as a science, and the tradition is particularly strong in subfields like demography (the statistical study of populations). That said, other social science fields tend to be much more quantitative in their approach—economics and political science, in particular. In this way, the field of anthropology is actually quite similar to sociology; indeed, many colleges and universities have departments of "anthropology and sociology," with their faculty drawn from both fields. The differences between the fields are historical and somewhat arbitrary, and both fields study human behavior and human societies. Generally speaking, anthropologists focus more on culture (and historically have tended to focus on non-Western cultures, although that has changed). Anthropologists also rely more exclusively on qualitative methods—particularly ethnography, which involves direct observation of a particular culture. These differences can be overstated, however, and perhaps the most important distinction between anthropologists and sociologists is that they speak to other scholars within their respective fields—which means their research builds upon particular traditions and perspectives. For example, if you compare Victor's sociological study of unemployment, *Cut Loose: Jobless and Hopeless in an Unfair Economy* (2015), with two excellent anthropological studies of unemployment—*End of the Line*, by Kathryn Marie Dudley (1997), and *A Company of One*, by Carrie M. Lane (2011)—you will find their approaches to be very different in terms of the theories they discuss and the scholars they cite, even though all three are qualitative studies.For our part, we reject a view that either quantitative or qualitative methods are superior. The qualitative versus quantitative debate is fueled by researchers who may prefer one approach over another, either because their own research questions are better suited to one approach or because they happened to have been trained in one method. In this text, we'll operate from the perspective that qualitative and quantitative methods are *complementary* rather than competing. The two approaches have different goals, strengths, and weaknesses, which we'll explore in later chapters.



If you compare the studies on unemployment referenced by sociologist Victor Tan Chen in his book *Cut Loose* (2015) to those cited by anthropologists Carrie M. Lane and Kathryn Marie Dudley in the books they wrote, you'll find some overlap, but lots of divergence. Every scientific field will highlight its own key theories and empirical research. *University of California Press* / *Cornell University Press* / *University of Chicago Press*

Another important thing to keep in mind about sociological research from the outset is that sociologists aim to explain *patterns* in society. Most of the time, a pattern will not explain every single person's experience—a fact about sociology that is both fascinating and frustrating. It is fascinating because even though the individuals who create a pattern

may not be the same over time and may not even know one another, collectively they create a pattern. Those new to sociology may find these patterns frustrating, however, because they may believe that the patterns that describe their gender, their age, or some other facet of their lives don't really represent their experience. It's true. A pattern can exist among your cohort without your individual participation in it.

Let's consider some specific examples. One area that sociologists commonly investigate is the impact of a person's *socioeconomic background* (the financial and other resources they were raised with) on their later life experiences and career prospects. A large body of research finds that whether a person comes from a poor or affluent family dramatically affects their educational attainment and achievement. For example, an influential study found that about a third of children from the richest 0.1 percent of U.S. families (those with annual incomes of $2.2 million or more) go to Ivy League or other elite colleges (Chetty et al. 2017). That proportion roughly matches the percentage of children from the poorest fifth of families (incomes of $25,000 or less) who go to *any* sort of college—four-year or two-year, public or private.

This finding—that family income shapes a child's educational experiences—is probably not that shocking to many of us, even if we know someone who may be an exception to the rule. Sometimes the patterns that social scientists observe fit our commonly held beliefs about the way the world works. When this happens, we don't tend to take issue with the fact that patterns don't necessarily represent all people's experiences. But what happens when the patterns disrupt our assumptions?



Seminal research by education scholars Myra and David Sadker (1994) described how gender bias creeps into the classroom. *Max Fischer, via Pexels*

For example, did you know that teachers are far more likely to encourage boys to think critically in class by asking them to expand on the answers they give and directly commenting on their remarks and observations? When girls speak up in class, teachers are more likely to simply nod and move on. The *pattern* of teachers engaging in more complex interactions with boys means that boys and girls do not receive the same educational experience in school (Sadker and

Sadker 1994).Some people will refute that such a pattern actually exists, citing evidence from their own experience. *They* knew of a teacher who paid equal attention to students regardless of their gender—or better yet, *they* knew a teacher who favored girls over boys. (Here, people are relying on anecdotal evidence and committing what is sometimes called the "person who" fallacy: *I know a person who …*) But objecting to a social pattern on the grounds that it doesn't match one's individual experience misses the point about patterns. You will find exceptions to virtually any social trend that you encounter. As sociologists who look at the big picture of social life, however, we are interested in that *overall* pattern—what happens on average, in the aggregate, on the whole. (Of course, we're also interested in how frequent or consistent the trend is—that is, just how many exceptions there are.)

Remember that as social scientists, you're going to be dealing with probabilities, not certainties. In this sense, social science can be a trickier and more complex endeavor than other approaches to learning you will come across in your academic or professional life.

Key Takeaways

1. Sociology, the study of social life and social groups, follows other social sciences in trying to identify patterns in society.
2. Although individuals make up patterns, every individual need not be a part of a pattern for that pattern to exist.
3. Qualitative methods are those that yield data such as words or pictures. Quantitative methods are those that yield data such as numbers. Sociologists use both qualitative and quantitative methods. While different, these methods are often complementary.

Exercise

Page or scroll through a few popular magazines or news sources. Pull out any examples you see of results from social science research being discussed. How much information about the research is provided? What questions do you have about the research? To what extent will the research shape your actions or beliefs? How, if at all, is your answer to that question based on your confidence in the research described?

1.3. Design and Goals of This Textbook



Like constructing a sturdy building, conducting a well-crafted research study takes time, lots of hard work, and sound planning. As later chapters will describe, before you begin your observations and analysis, you need to build a strong foundation in the existing research literature, situating your creative work in relation to evolving theoretical understandings and previous empirical results. *Josue Isai Ramos Figueroa, via Unsplash*

The goal of this textbook, *The Craft of Sociological Research*, is to get you a concise but well-rounded understanding of how sociologists conduct research. We call it a "craft" because, much like a practical art such as carpentry or needlework, doing it exceptionally well requires a balanced blend of creativity, rigor, resourcefulness, and passion. Its practitioners are constantly learning and improving their methodological skills, and like other crafts, it encompasses a vast range of styles and approaches.

Unlike most methods textbooks, we start by talking about the social context of research, rather than leaving it for the end of the book. In **Chapter 2: Using Sociology in Everyday Life**, we discuss the different approaches to sociological research—academic, applied, and public. The idea of covering this material early is to help you understand the actual point of conducting research as well as the possibilities it offers for your own life and career. That way, you can read the rest of the textbook in a more focused way, with these goals and aspirations in mind.

The next several chapters outline the general procedures involved in planning a research project—starting with big-picture considerations and moving to more concrete decisions. We've set up the textbook this way so that it provides students with a blueprint for designing their own study. Your instructor may have you write up a research proposal as a final paper for this class, or you may plan to write a thesis. The material in this book is structured so that you can get started right away in thinking about and planning this potential project. With that goal in mind, we set aside material directly related to writing up a study or research proposal and put it in **Appendix A: Presenting, Writing, and Publishing**. That way, you can read it at any point when it's useful for you to start thinking about how to organize your research-related writing and presentations.

In **Chapter 3: The Role of Theory in Research**, we describe the various perspectives that sociologists bring to the study of social life. We'll show you how theory guides the ways that we collect and analyze data, and theory, in turn, emerges from the data we gather. As we describe in detail, our theories and empirical work always build upon each other. In **Chapter 4: Research Questions**, we discuss how to generate interesting and feasible questions that you could answer with research. In **Chapter 5: Research Design**, we cover the process for conducting a literature review and thereby identifying a research problem that justifies your proposed study.

The next several chapters go over general considerations that pertain to sociologists using any research method. In **Chapter 6: Sampling**, we'll review strategies for finding a good sample for your research—whether you are trying to find a random sample that is truly representative of your population of interest, or seeking a purposive sample that fits the needs of a particular theory. In **Chapter 7: Measuring the Social World**, we'll discuss how to identify key concepts of interest for your study and how to measure them in the real world. In **Chapter 8: Ethics**, we'll talk about the moral quandaries that researchers often have to confront, and how best to conduct your research with both ethical and practical considerations in mind.

We'll then move on to the most exciting part of the research process: collecting and analyzing data. We give equal weight to both qualitative and quantitative methods, and we take pains to show how these two approaches complement each other. We'll start with qualitative approaches, as covered in **Chapter 9: Ethnography**,**Chapter 10: In-Depth Interviews**, and **Chapter 11: Qualitative Data Analysis**. In **Chapter 12: Experiments**, we'll talk about the classic scientific method of conducting deductive analysis using experimental data. Then we'll turn to quantitative methods that typically employ this deductive approach while using nonexperimental data. **Chapter 13: Surveys** goes over how to collect quantitative data through the fielding of surveys, and **Chapter 14: Quantitative Data Analysis** provides an entry-level primer on how to interpret the data you collect statistically. Finally, we'll briefly cover content analysis, historical sociological research, and social network analysis in **Chapter 15: Materials-Based Methods**. While many more methods exist for studying social life, we feel that these chapters cover the methods most commonly used in sociological research.

What will be the payoff to you for reading all this material? Hopefully, you will feel you've gained a real understanding of sociological research methods, how and why they are relevant to you, and how they can help us make sense of our world. As with other crafts, mastering sociological research takes a great deal of time and effort, but it will transform the way you think about society and, perhaps, give you some powerful tools to change it.

1.4. References

Chen, Victor Tan. 2015. *Cut Loose: Jobless and Hopeless in an Unfair Economy*. Oakland: University of California Press.

Cherlin, Andrew J. 2014. *Labor's Love Lost: The Rise and Fall of the Working-Class Family in America.* New York: Russell Sage Foundation.

Chetty, Raj, John Friedman, Emmanuel Saez, Nicholas Turner, and Danny Yagan. 2017. "Mobility Report Cards: The Role of Colleges in Intergenerational Mobility." Working Paper No. 23618, National Bureau of Economic Research, Cambridge, MA. doi: 10.3386/w23618.

Coontz, Stephanie. 1992. *The Way We Never Were: American Families and the Nostalgia Trap*. New York: Basic Books.

Dudley, Kathryn Marie. 1997. *The End of the Line: Lost Jobs, New Lives in Postindustrial America*. Chicago: University of Chicago Press.

Edin, Kathryn, and Timothy J. Nelson. 2013. *Doing the Best I Can: Fatherhood in the Inner City*. Berkeley: University of California Press.

Handelsman, J. B. 1986. "I'm a Social Scientist, Michael." *New Yorker*, April 27.

Hernandez, Donald J. 1993. *America's Children: Resources from Family, Government, and the Economy*. New York: Russell Sage Foundation.

Hochschild, Arlie Russell. 2012. *The Managed Heart: Commercialization of Human Feeling, Updated with a New Preface*. 3rd ed. Berkeley: University of California Press.

Lane, Carrie M. 2011. A *Company of One: Insecurity, Independence, and the New World of White-Collar Unemployment*. Ithaca, NY: Cornell University Press.

Lazarus, Richard S., and Susan Folkman. 1984. *Stress, Appraisal, and Coping*. New York: Springer.

Merton, Robert King. 1957. *Social Theory and Social Structure*. Glencoe, IL: Free Press.

Mikkelson, Barbara. 2005. "Grandma's Cooking Secret." *Snopes*. Retrieved October 5, 2021 (https://www.snopes.com/fact-check/grandmas-cooking-secret).

Rogers, Simon. 2012. "Data Are or Data Is?" *Guardian*, July 8.

Sadker, Myra Pollack, and David Miller Sadker. 1994. *Failing at Fairness: How America's Schools Cheat Girls*. New York: Scribner's.

Turnbull, H. W., ed. 1959. *The Correspondence of Isaac Newton*. Vol. 1. Cambridge, UK: Royal Society, University Press.

Tyson, Neil deGrasse. 2016. "In Science, When Human Behavior Enters the Equation, Things Go Nonlinear." *@neiltyson*. Retrieved October 20, 2021 (https://x.com/neiltyson/status/695759776752496640).

2. USING SOCIOLOGY IN EVERYDAY LIFE



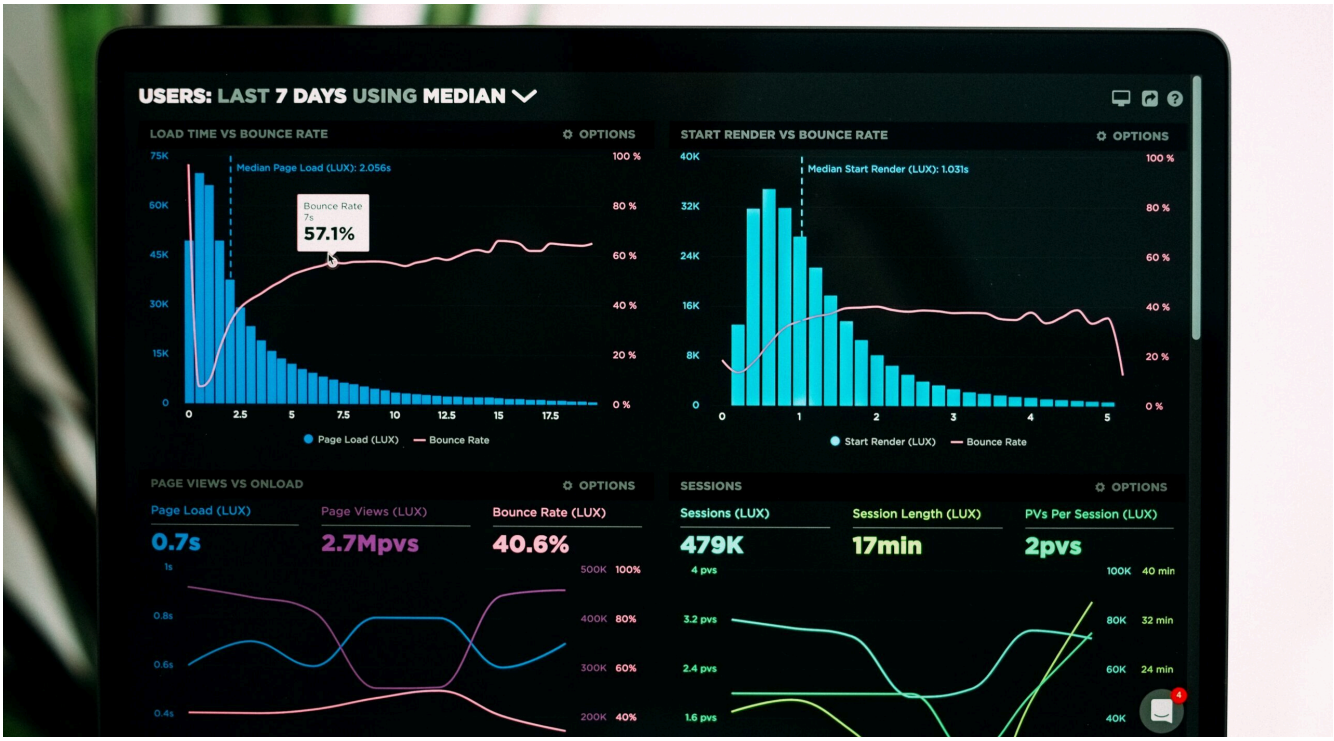
At this point, you may be wondering about whether research methods actually have any relevance to your own life. Whether or not you choose to work as a sociologist, we believe you should care about research methods for two reasons. First, research methods are regularly applied to solve social problems and issues that shape how societies are organized—which means you have to live with the results of research methods every day of your life. Second, understanding research methods might help you find a meaningful career and make a difference in the society you live in.

We'll return to these two main uses of sociology throughout this chapter. Before we dive in, however, we should make a distinction between types of science. Generally speaking, we can classify research across the many fields of natural and social sciences (including sociology) based on the end purpose. **Basic research**, also called *pure* science, is research for its own sake. In this case, scientists are motivated to conduct research simply because they happen to be interested in a topic and want to contribute to a broader understanding of it. **Applied research** brings scientific knowledge to bear on a practical problem. It is often client-focused, meaning that researchers are investigating questions posed by someone other than themselves—often, a group that is paying them. Although many scientific disciplines do a mix of basic and applied research, some are more geared toward one side of this continuum. Engineering, for example, uses knowledge from physics and chemistry to build stronger bridges or design more efficient engines. Medicine uses biological knowledge to treat human ailments. Businesses tend to focus more on applied research, given its practical value, while universities conduct both basic and applied research.

For their part, sociologists tend to focus more on basic research than other social scientists do, though plenty of researchers trained in sociology wind up doing applied work, as we will see. They may not necessarily identify themselves as "sociologists," but they were taught sociological research skills and approach their work from a sociological perspective. These researchers are frequently hired to conduct **evaluation research**, helping assess whether social programs—anything from local domestic violence shelters to federal government assistance programs—are effective in pursuing their missions. Many sociologists also work for **think tanks**, which are private or government-run organizations that conduct research to help governments, corporations, activists, or other groups make decisions about the best policies to pursue or reforms to push forward. Even when people trained as sociologists

decide to pursue teaching and researching careers within academia, they may be interested in applying their sociological knowledge more broadly—that is, making their research findings accessible to the larger world through practices of **public sociology**.

In recent years, the demand for skilled social scientific researchers has grown. As advances in digital technologies make it easier and easier to collect information about users of websites, apps, and networked devices, they generate a complex and unstructured mass of data—what is known as **big data**—that is waiting to be analyzed by someone. Sociologists often get jobs in **data science**, an interdisciplinary field that uses computer coding and social scientific analysis methods to address the data-related problems facing corporations, governments, and other organizations.



Sociology students with more advanced quantitative skills can pursue careers in data science, working on behalf of companies and agencies to bring order and meaning to the endless cascades of info generated by today's digital culture. *Luke Chesser, via Unsplash*

Here's a list of the types of jobs that people trained as sociologists typically work:

- teaching at the high school or university level
- studying social networks for social media companies
- doing research for advocacy organizations and think tanks
- working for the U.S. Census Bureau and other government agencies
- conducting market research in the advertising industry
- writing or reviewing grant applications
- tracking social well-being and other patient outcomes for hospitals, pharmaceutical companies, and other healthcare organizations
- providing therapy or counseling to individuals or groups
- facilitating groups and mediating conflicts, sometimes on behalf of courts

- working in human resources or otherwise using sociological knowledge of organizations and social networks in a corporate management role
- helping with the collection of data for large-scale social science studies such as the General Social Survey

Understanding research methods is important in all these jobs, and learning these skills can lead you to a satisfying and socially meaningful career.

In this chapter, we start off by talking about the sorts of basic research that academic sociologists—sociologists based at universities and colleges—typically conduct. Then we'll discuss the numerous ways that sociological research matters in the world outside the academy—not only in terms of applied sociology, but also in broader terms. As you will see, the skills that good researchers possess are highly valued both within the **public sector** (government bodies, such as federal agencies and state legislatures) and within the **private sector** (companies, nonprofit organizations, and the like). Sociological research offers a number of career routes you can take, both academic and professional, and sociologists can make use of research to support efforts to initiate and sustain social change—as policymakers and activists, or simply as citizens and consumers.

Opening chapter image credit: *Sachith Ravishka Kodikara, via Pexels. Adapted by Bizhan Khodabandeh.*

2.1. Sociologists as Scholars



Sociologists who focus on basic research typically teach at colleges and universities, where in most cases they must juggle their roles as both scholars and instructors—even if their research output tends to be more decisive in their career advancement. (If you've ever had professors who were just not into teaching, this may be part of the reason.) *ICSA, via Pexels*

Learning Objectives

1. Distinguish among the various types of sociologists who focus on basic research and the various institutions that employ them.
2. Understand academic tenure and how the decline in the number of faculty members who have it has changed the workforce at colleges and universities.

When we think of sociologists who conduct basic research, we're largely talking about sociologists at higher education institutions who engage in both teaching and research (see the sidebar **Researchers and Research Communities**). Sociology departments differ greatly in the types of sociology they specialize in. A particular department often has strengths in one or a few subfields of sociology—say, medical sociology or the sociology of race and ethnicity. These specialties can be part of a long tradition for that department, perhaps because a subfield's most prominent scholars are or were based at that department. Sociology departments may also be multidisciplinary in that they connect their work with another discipline, like criminology or public health—perhaps teaching classes to students in both fields, running a center or institute with scholars from both fields, or even explicitly bridging two disciplines (for instance, by identifying themselves as a Department of Sociology *and* Criminology). Many academic sociologists also wind up

working outside of sociology departments. If they have a background in organizational sociology, for instance, they may become faculty members at business schools, where they teach students about best practices in managing companies. If they study health outcomes across populations, they may be employed at medical schools or schools of public health. Other sociologists teach at schools of public policy, education, and criminal justice, which hire scholars who do work in related subject areas.

Even if a school's sociology department doesn't specialize in a specific subfield, its faculty most certainly do. They will teach and conduct research in their chosen areas of specialty. Through their academic writing and lecturing, they will regularly interact with other scholars at their home institution or elsewhere who share those interests. **Professional associations** are a critical part of that networking—for both faculty and students. A wide range of sociology-related professional associations put on annual conferences and publish journals focused on their areas of interest. At the national and international level, there are professional associations representing the entire field, such as the American Sociological Association, the Canadian Sociological Association, the European Sociological Association, the Korean Sociological Association, the International Sociological Association, and the Latin American Association of Sociology. These associations put on annual meetings in major cities that attract thousands of scholars, who attend panels devoted to specific research or teaching themes and otherwise network at meals and social events. In addition to their conferences, the larger associations will have *sections* or *networks* that specialize in particular subfields, giving sociologists access to smaller communities of like-minded peers both during and after the group's annual meetings.



Sociologists give talks about their research at academic conferences, where other scholars in attendance will pose questions and provide feedback. These conferences are usually sponsored by professional associations—such as the Society for the Advancement of Socio-Economics (SASE), an international group of scholars who study economic sociology and political economy. The picture above was taken at a panel session at a SASE conference in Amsterdam.

While it is a good idea to attend the annual meetings of these larger groups, students may find regional professional associations to be more welcoming and less bewildering (see **Appendix A: Presenting, Writing, and Publishing** for a primer on presenting at conferences). In the United States, regional conferences are sponsored by groups like the Eastern Sociological Society, the Southern Sociological Society, and the Pacific Sociological Association. Sociologists may eschew these catchall associations and instead attend conferences for groups devoted to their research areas, such as the Rural Sociological Society and the Society for the Study of Social Problems (the latter being a prominent sociological association whose mission is to further social justice). In fact, some sociologists intentionally seek out multidisciplinary associations, like the Academy of Management (a group focused on organizational studies that is largely attended by sociologists and other scholars based at business schools) or the Society for the Advancement of Socio-Economics (which attracts economic sociologists and political economists, among others).

If you pursue a career as an academic sociologist, you will inevitably find yourselves a part of these intellectual conversations and debates within your areas of interest. Sometimes, the life of an academic can be glamorous—being invited to give talks at other universities, attending conferences around the globe. However, as we describe in the sidebar **Researchers and Research Communities**, academic livelihoods can be precarious, and the competition is often intense across the range of available faculty positions. To have a shot at a tenure-track position, you usually need to prioritize research over teaching—a fact that many students with hopes to teach at the college level may not realize. You typically will want to obtain a PhD to get stable employment, and it is generally a good idea to study the field that you will wind up teaching in—departments of sociology tend to prefer hiring people with PhDs in sociology, for example.

Researchers and Research Communities

Sometimes nonscientists are confused about the various actors and groups involved in conducting scientific research. Not all of them are professors. Here is a breakdown of the various terms used to describe researchers and research communities—as seen in the field of sociology, but more broadly, too.

Doctorates and other degrees. To teach sociology (or any academic discipline, for that matter) at a four-year university, instructors in many countries typically must have an advanced degree known as a *doctorate* or *PhD* (in Latin, *philosophiae doctor*, or "doctor of philosophy"). This requires the scholar to spend several years taking classes and then several more years conducting an independent and original research project known as a **dissertation**, or *doctoral thesis*. Oftentimes, the scholar gets a master's degree after two or so years in graduate school, perhaps after completing a *master's thesis*—an independent project on a smaller scale than a dissertation. They may stop their studies at that point or continue to complete their PhD. For dissertations and theses, students typically need to get three or more scholars from their department (and frequently from outside the department or school as well) to serve on a *thesis committee*. Members of the committee review the student's work and must all sign off on both the proposal for the study (the thesis *prospectus*) and the completed study for students to receive their degree. One faculty member, usually the student's main advisor, chairs that committee. In most departments, students present their prospectus and completed study to their committee and respond to the committee members' feedback and criticisms in a storied oral ritual known as the dissertation or thesis *defense*—which can sometimes be as brutal as this relevant XKCD comic suggests (but usually not, thankfully).



Before they receive their doctorates, PhD candidates must give a presentation about their dissertation projects and answer questions from faculty as part of a nerve-racking but typically nonviolent session known as a thesis defense. *Randall Munroe, via* XKCD

Professors, faculty members, departments. After scholars earn their PhD, they frequently go on to work as professors. Within a college or university, these scholars will be identified as *faculty members* of particular *departments*, which are generally divided along disciplinary lines (e.g., the Department of Sociology). U.S. universities distinguish between *full professors* (typically just called "professors" in their job titles), *associate professors*, and *assistant professors*. One key distinction between these types of professors is whether they have **academic tenure**, a form of job security that makes it difficult to fire a professor. Tenure is meant to allow professors to study and write about whatever they want without fear of political repercussions for speaking the "truth." A college or university will grant a professor tenure after a lengthy process of tenure review, which involves other faculty at that school and other schools being asked to assess the quality and quantity of the professor's research, teaching, and other contributions. Typically, associate and full professors (also known as *senior faculty*) have tenure, while assistant professors (also known as *junior faculty* or *tenure-track faculty*) are in the process of applying for it. Assistant professors typically "go up for tenure" after spending six or so years as junior faculty.

Instructors, lecturers, teaching faculty, adjunct faculty. The number of tenured and tenure-track faculty (collectively known as *tenure-line faculty*) in the United States has declined dramatically in recent decades, in part because of desires by universities to hire cheaper instructors without tenure protections. According to data compiled by the American Association of University Professors (AAUP), an advocacy organization representing academics, the numbers of "contingent faculty"—part-time- and full-time faculty appointed off the tenure track—has risen dramatically over the past several decades: from 43 percent of the academic labor force in 1975, to 69 percent in 2019 (AAUP 2020).

In the United States, faculty who are not tenured and not on the tenure track are often referred to *teaching faculty* and may have the title *instructor* or *lecturer* (to distinguish them from professors). Teaching faculty generally are not required to conduct research in addition to their teaching duties, but many do substantial amounts of research—perhaps in the hope of landing one of the dwindling number of available tenure-track positions. Sometimes teaching faculty are full-time and may even have the rank of "professor," but this is typically qualified with the term "teaching professor" to underscore that they do not primarily conduct research.

Full-time non-tenure-track faculty accounted for 18 percent of the academic labor force in 2020, nearly double their proportion in 1975, but the numbers of *part-time* faculty have grown even more—to 52 percent of the labor force in 2019. Frequently, part-time instructors or lecturers do not have long-term contracts with the university, but are paid for each class they teach. These teachers are called **adjunct faculty**, or *adjuncts*, and they tend to suffer from low wages and job insecurity. The AAUP describes their pay rates as "appallingly low," calculating that on average they received just $3,500 per course section taught in the 2019–20 academic year, with less than a third having employers who offered health insurance (AAUP 2020).

Research universities, liberal arts colleges, and community colleges. Some academic institutions focus primarily on research, while others focus primarily on teaching. In the United States, an influential ranking system, the Carnegie Classification of Institutions of Higher Education, distinguishes colleges and universities based on how much research their scholars do. At the research-intensive end are so-called **R1 schools** that engage in "very high research activity." Notably, these schools have large populations of doctoral (PhD) students who are training to be professors. Many R1 schools are Ivy League schools with large endowments, but they also include many large public universities—particularly so-called *land-grant universities* (like Virginia Tech) largely created through legislation passed in the nineteenth century that granted federal lands for the purpose of founding universities.

Our school, Virginia Commonwealth University, is considered an R1 institution, though we should note that this designation can depend on particular departments or schools with a large research presence (like our medical school), and other departments or schools within that university may not be so research-focused. Overall, however, the tenure-line faculty at research universities (whether R1 or lower-ranked) tend to spend more of their time on conducting research, often leaving much of the actual work of teaching to graduate students or teaching faculty. *Liberal arts colleges* tend to conduct less research than research universities. These schools have few or no graduate students, offer smaller classes, and have faculty who are primarily focused on their teaching duties. Even if they do not have graduate students, however, liberal arts colleges offer four-year (bachelor's) degrees. *Community colleges*, by contrast, offer only two-year (associate's) degrees. Faculty at community colleges often do not conduct research, and they do not always have PhDs—although given the competitive job market we discussed earlier, PhD graduates increasingly are winding up teaching at these schools.

Think-tank scholars, policymakers, independent scholars. Sometimes, social scientists work as researchers at think tanks, which are organizations devoted to conducting research. Think tanks can be government agencies, but many are private groups funded by governments, corporations, foundations, or wealthy donors. Unlike academics, think-tank scholars typically do not teach students, though many researchers have faculty appointments at universities in addition to being paid by think tanks. For most think tanks, the goal of their research is to influence **policymakers**—top government, corporate, or other officials who can create laws, regulations, rules, and other sorts of policies that affect people's behavior. As a result, much of the research that think tanks engage in is applied. That said, think-tank researchers are different from other nonacademic researchers in that they may do more in the way of basic research, and their paths often cross with those of academics. In fact, scholars at these institutions sometimes publish their work in academic journals and attend academic conferences. Their counterparts based at universities, in turn, will frequently cite and build on their work—a sign that they are part of the larger scientific conversation about basic research.

You may also find **independent scholars**—researchers without any affiliations to a university, think tank, or other institution—engaged in these sorts of debates about basic science. This is quite rare, however. In sociology (and across scientific fields, for that matter), your institutional credentials matter a great deal in how seriously your research is taken—for better, or for worse.

Key Takeaways

1. Sociologists specialize in particular subfields of sociology, which means that departments of sociology at different universities or colleges may focus on particular research areas, and sociologists often work in interdisciplinary departments or schools.
2. Outside their academic institutions, sociologists connect with each other through professional associations, which often focus on specific subfields or interdisciplinary studies.
3. Like all faculty at universities and colleges, academic sociologists are divided into tenured faculty, tenure-track faculty, and teaching faculty. The ranks of teaching faculty have grown enormously over recent decades, and many of these instructors have little in the way of income or job security.
4. Academic institutions differ greatly in terms of the mix of research and teaching they do, with R1 universities being the most research-intensive, and liberal arts colleges and community colleges focused more on teaching.
5. Sociologists also work outside academic institutions, with scholars at think tanks being the most engaged in basic research that is similar to what academic sociologists do.

2.2. Sociologists as Market Researchers



Market researchers gather their data using many of the research methods that sociologists rely on, such as surveys and interviews. Thus, the many skills that you gain studying sociology can be helpful for getting people to buy stuff, and not just asking why they buy so much stuff. *Christina Morillo, via Pexels*

Learning Objective

Describe the various types of research that sociologists conduct for marketing or other purposes on behalf of corporations.

On the opposite end of the spectrum from academic sociology is the applied work that sociologists do for companies. Many sociologists work in the private sector. Organizational sociology is a huge part of the curriculum taught at business schools, and sociological knowledge about how organizations and networks operate is useful in the specialized fields of human resources and organizational consulting (where sociologists advise companies or nonprofits about how to structure and run their workplaces) as well as in a variety of other management-related roles.

The type of applied research most frequently conducted in the private sector is **market research**. Market research is not itself a research method. Rather, it uses various methods to help guide companies in making decisions about increasing their sales or improving their performance.

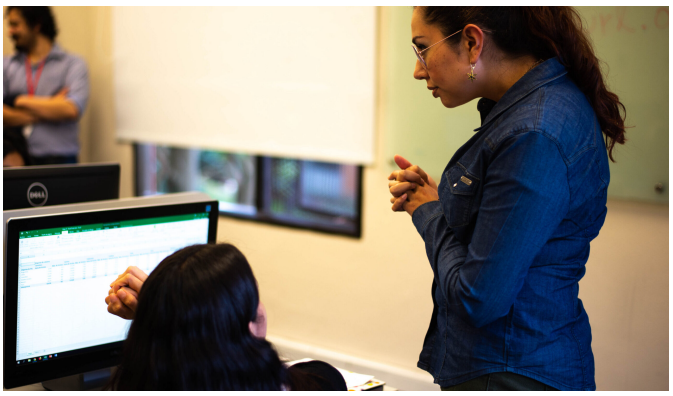
Let's say a firm wants to promote a particular product or service. Before they dive into a marketing campaign or invest money into research and development, they need to collect data about what their potential customers are looking for and what their potential competitors are doing. Oftentimes, they will want to get a better understanding of the

overall industry or local market that they're operating within, especially if they are moving into a new area of business. Companies can hire market research firms to do this data collection and analysis for them, or they may have in-house research operations.

When reaching out to customers, companies typically seek out particular niches—say, young, single college-educated consumers with plenty of disposable income. To appeal to this target demographic, it is important for a company to understand that population's preferences and tastes. That's where market researchers come in. Their methods can focus broadly on people's general attitudes, or more narrowly on their interactions with products and advertising. For example, market researchers routinely observe clientele in stores to watch which displays draw them in, and which they ignore. They conduct **focus groups** (an approach we'll examine further in **Chapter 10: In-Depth Interviewing**), where they interview a group of people in a particular target demographic and try to gain a fine-grained understanding of what issues they have that a particular product or service might address, or how they feel about existing market offerings. They also administer surveys to assess customers' use of a particular product or service and their satisfaction with it.

When companies track consumers' attitudes and behaviors to get a sense of the current and future state of a market, the research they do often looks a lot like the analysis of social and cultural trends that academic sociologists regularly do. Indeed, well-known global research firms such as Nielsen, Gallup, and Ipsos often hire trained survey researchers with social science backgrounds. Nielsen is a global market research firm that is best known for its Nielsen ratings, a measurement system that calculates the size of audiences for television and many other media markets. Gallup and Ipsos are well-regarded among social scientists for their public opinion polls (which aside from their market value, can be useful for government officials making decisions about policy, as we will describe later). But both companies also offer analytics services, consulting with government and corporate clients around the globe. (**Data analytics** involves the use of scientific methods to interpret available information, identify trends and patterns, and map complex relationships.)

As you might imagine, tech companies like Apple and Google regularly generate mountains of data about their users, and they need skilled researchers to make sense of that data. For that reason, sociologists with strong quantitative skills often work for these companies as *data scientists*, exploring what useful patterns emerge in the big data they have collected. Social media companies in particular are in need of researchers who understand **social network analysis**—the study of social ties, how they emerge and evolve, and what consequences they have (we'll discuss social network analysis in **Chapter 15: Materials-Based Methods**).



To pursue a career in data science, you might want to take undergraduate or graduate classes in higher-level quantitative methods, which are taught in sociology and statistics departments and elsewhere—such as this data science class offered by the communications department at Politécnico Grancolombiano in Bogotá, Colombia. *Departamento de Comunicaciones del Politécnico Grancolombiano, via Flickr*

While companies often prize researchers with the advanced quantitative skills needed to analyze big data, there is also a growing appreciation in the private sector for what the sociologist Tricia Wang (2016) has called "thick data"—ethnographic research that generates rich and nuanced qualitative data that can help companies adapt the ways they think about and measure a particular problem. (We'll have more to say about **ethnography** and the sort of **thick description** it prioritizes in **Chapter 9: Ethnography**.) Market research can employ both qualitative and quantitative methods, as we can see most vividly in the burgeoning field of market research called **user experience research**, or *design research*. The goal of UX research is to better understand the perspective of the "end user"—the person who ultimately will use the company's product or service. Through conversations

with potential or current customers and observations of their interactions with that product or service, UX researchers gather data that can be used to improve its design.

Companies who conduct UX research regularly hire sociologists: in-depth interviewers who are good at probing users' motivations and needs, ethnographers who can parse out the various ways they use the product, and survey and experimental researchers who can obtain quantitative measures of their behaviors and attitudes. Prominent UX firms in the United States include Bare International, a global company headquartered in Fairfax, Virginia, and California-based AnswerLab, both of which conduct research for a wide variety of corporate clients. Large companies like Facebook also employ their own UX researchers—qualitative and quantitative—and even offer research internships.

In the corporate sector, you don't need an advanced degree to get a job conducting market research. Internships are perhaps the best way to break into these commercial fields, though sometimes a master's degree or PhD will be decisive for more technically demanding research positions such as those involving data science. Indeed, with the growth of big data, graduate programs have proliferated in recent years to train data scientists, who have job prospects that extend well beyond the private sector. (Our own sociology department offers such a master's program, which includes coursework in digital sociological research methods.) In particular, government agencies and think tanks have a great need for social scientists to crunch their numbers, as we will describe next.

Key Takeaways

1. Sociologists who work in the corporate sector typically conduct market research to improve a company's sales and performance.
2. With the growth of big data, demand is growing for data scientists who can use advanced quantitative methods to understand large amounts of unstructured data.
3. Qualitative sociologists can also use their skills in conducting in-depth interviews, focus groups, and qualitative analysis to contribute to market research, particularly in the area of user experience (UX) research.

2.3. Sociologists as Policy Advisors and Evaluators

Learning Objectives

1. Describe the various roles that sociologists can perform to evaluate policies and advise policymakers.
2. Distinguish between the policy work that sociologists can do in the public sector and the private sector.

Sociologists who work for universities and corporations rarely talk to one another, given the starkly different goals of their research. However, one important space where sociologists who conduct basic and applied research interact is the arena of public policy. Here, we can see scientific ideas and findings dramatically shaping the social structure that determines how people live their lives.

Influencing policy is a direct way that social scientists can have an impact on real people. Corporations, governments, and other organizations each have their own *policies*—rules and regulations for how things should be done. When social scientists talk about changing policies, however, they're typically talking about changing government policies, or *public policies*, given that governments typically have the greater authority to enact and implement policies with a broad scope and reach. Of particular interest to sociologists is **social policy**, a type of public policy that has to do with human well-being—from health care to education to labor regulations.[1]

1. Harvard University's Malcolm Wiener Center for Social Policy defines social policy as "public policy and practice in the areas of health care, human services, criminal justice, inequality, education, and labor."



Sociologist Daniel Patrick Moynihan (1927–2003) worked for four decades in government, including as an influential U.S. senator representing New York. President Ronald Reagan (1911–2004) and First Lady Michelle Obama (1964–) studied sociology in college. *John Mathew Smith (Moynihan) and the White House (Reagan and Obama), via Wikimedia Commons*

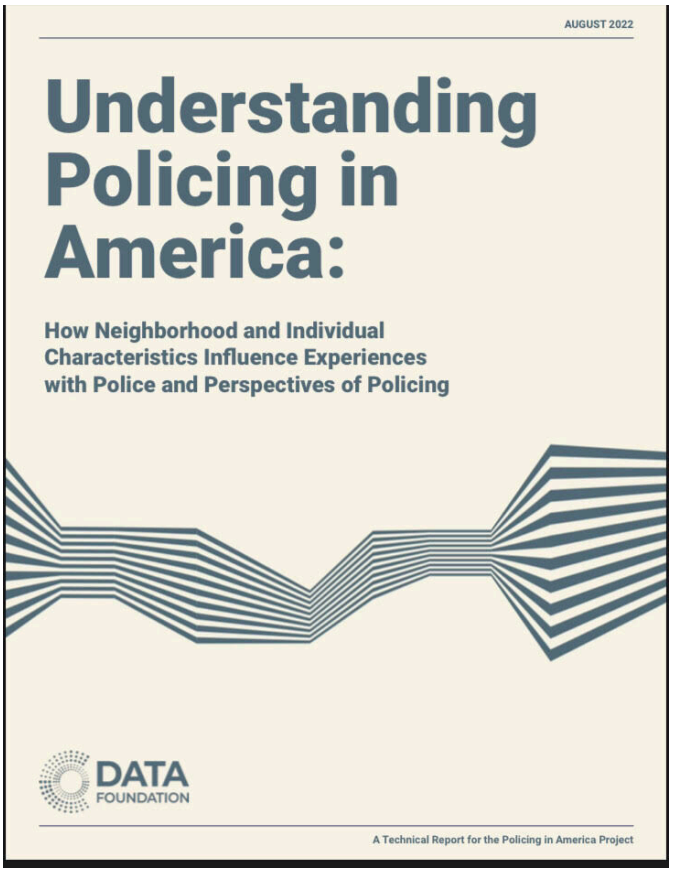
Sometimes sociologists make policy directly—either by being elected to office or appointed to top positions where they can actually determine the laws or regulations on the books. One of the most prominent U.S. sociologists who became an incredibly influential policymaker was Daniel Patrick Moynihan, who served in a variety of key executive-branch positions for both Democratic and Republican administrations and represented New York in the Senate for decades. Political leaders like Ronald Reagan, the U.S. president, and Michelle Obama, the first African American First Lady, never worked as professional sociologists but majored in sociology in college. (If you're looking for other big names, the ASA maintains a list of sociology majors who went on to fame in politics and other fields.)

For the most part, however, sociologists assist policymakers as advisors. They help provide services, by researching what particular communities need and how to deliver those services effectively. They measure the performance of government agencies, and they collect and assess the evidence in favor of or against certain policy reforms. Sociologists based at academic institutions sometimes do this work as consultants for government agencies, but larger agencies typically hire their own researchers.

Organizations outside of the government also generate scientific knowledge that can influence policy. We have already discussed think tanks, whose scholars are often engaged in scientific debates. As we will describe later in this section, there is a much broader array of policy-oriented research settings, many of which conduct research for practical purposes.

In their work on policy matters, social scientists will frequently draft **policy briefs**, which summarize a particular issue and the recommended policy approaches for dealing with it. The idea is that policymakers can quickly read these briefs (which often break down the existing scientific literature in a concise and accessible way) and then make informed decisions about what laws, regulations, or procedures to put in place. An organization's researchers may also write a **white paper**—a report that is similar to a policy brief but typically meant to present an organization's take on the issue and spark a larger discussion and debate. A white paper may be geared toward policymakers, the public, or a variety of

other audiences. (Companies can even produce white papers—generally, to convince customers or other stakeholders about the merits of a particular product or business strategy.)



Think tanks often produce policy briefs, white papers, technical reports, and other publications for policymakers or the public, such as this report (Williams et al. 2022) on public perceptions of the police that features a sociologist's work.

In terms of original research, social scientists working on policy matters often conduct what is called evaluation research, the analysis of the real-world consequences of social programs or policies. This type of research is frequently deployed when some form of **policy intervention** is planned—that is, government officials want to launch or expand particular programs or enact or alter particular laws or regulations. These policy interventions may involve sweeping reforms, such as efforts to overhaul a country's health or social welfare systems. However, they can also entail smaller-scale changes, such as revising a school district's curricula or deciding whether to continue a support program for at-risk teenagers.

As the name suggests, evaluation research often involves researchers finding ways to measure whether the policy intervention in question has had its intended consequences. It can also be more exploratory, aimed at defining or diagnosing a particular social problem to determine whether an intervention may be warranted. The fact that evaluation research touches on both social problems and policy solutions makes it a natural fit for sociologists.

Broadly speaking, a **program** is made up of different interventions that individuals or groups receive—anything from giving low-income households access to housing vouchers, to enrolling individuals in a smoking-cessation class. The evaluation of a program usually starts with a **needs assessment**. This step entails systematic efforts to determine what assistance may be needed by the individual clients or communities that the program serves. Typically, a needs assessment is designed to compare what the actual situation is with what an appropriate, acceptable, or ideal standard would be. A relevant program may then be designed or restructured to address any important gaps between that standard and the current reality. You'll notice that the sort of research being done in program evaluation is somewhat similar to the market research that corporations do, but the target demographic is not a for-profit firm's customers, but rather a government agency or nonprofit organization's clients. Like market researchers, program evaluators can be hired in house (by the agency or organization that is evaluating its own programs), or they can work on a contract basis, as organizational or community consultants.

An **outcomes assessment** is another common step in the program evaluation process. Here, researchers determine whether a particular program has achieved its intended outcomes. They often start by identifying and measuring the program's "inputs" and "outputs." *Inputs* are the various resources the program took advantage of—its funding, of course, but also less obvious resources such as the organization's physical location, the equipment it owns, and the experience and knowledge of its staff. *Outputs* are the tangible results of the program. These typically include quantitative measures, such as the number of clients helped or the number of staff members trained to implement an

intervention, but they might also include qualitative data, such as feedback from interviewed clients. *Outcomes* are the observed changes, whether intended or unintended, that occurred as a result of the program.

After measuring these outcomes, researchers need to determine if the program actually achieved its desired results (Wagenaar 2002). If they have been rigorously trained as social scientists, they will be careful to rule out alternative explanations for the success or failure of a program, rather than just taking, say, higher numbers of clients helped or staff trained as proof. Nowadays, many program evaluations use the approach of a randomized controlled trial to provide harder evidence that programs actually work (we'll discuss this methodological approach further in **Chapter 12: Experiments**).

While some sociology programs at the college or university level teach program evaluation, students interested in pursuing a career in this area typically need to learn these specialized skills by doing internships or further studies in fields more explicitly focused on policy analysis. For example, undergraduate sociology majors often end up applying to master's programs in public administration, public policy, public health, or data science to be trained in the approaches used in particular sectors. That said, having a sharp understanding of good research design is prized across policy circles—both within government agencies and outside them, as we will discuss in turn.

Government Research



The U.S. Census Bureau (whose headquarters in Suitland, Maryland, are pictured above) is mandated by law to count every person in the country every 10 years, with the results deciding everything from the number of representatives each state has in Congress to the amount of federal funding that localities receive. In addition to implementing the decennial census, the bureau also collects a dizzying array of statistics about U.S. households based on its estimates from survey data. *U.S. Census Bureau, via Wikimedia Commons*

In the United States, the federal government is one of the largest employers of applied sociological researchers. Several

agencies stand out in particular. Given how much sociologists rely on the Census Bureau's data for their quantitative analyses, it's not surprising that the bureau hires many sociologists—particularly, those with training in **demography**, a sociological field that focuses on the statistical study of populations. For a similar reason, the Bureau of Labor Statistics, which collects data on workers and employers, is a good home for sociologists with strong quantitative skills. Those two bureaus are housed at the Department of Commerce and Department of Labor, respectively, which employ social scientists at various levels. Likewise, the Department of Health and Human Services and the Department of Housing and Urban Development—two cabinet-level agencies that oversee a wide range of social policies for the federal government—regularly hire social scientists to evaluate those policies and perform other policy-oriented tasks. On Capitol Hill, sociologists with policy knowledge sometimes are hired as legislative aides or otherwise support or advise members of Congress.

A number of government agencies are specifically tasked with the goal of conducting research to keep policymakers informed. These agencies can be categorized as *governmental* think tanks. An example in the United States would be the Congressional Research Service—known as "Congress's think tank"—which provides nonpartisan and oftentimes confidential advice to lawmakers about the impacts of proposed or enacted legislation and other regulatory matters.

Specializing in particular sociological methods can go a long way in landing you a job at any one of these agencies. In particular, statistical analysis skills are prized in governmental circles, given how much policymakers tend to rely on quantitative data in making decisions about where to allocate resources. Knowledge of substantive areas of research can also be helpful. For example, sociologists with a focus on criminology work for the Department of Justice, rural sociologists work for the Department of Agriculture and the Department of the Interior, and sociologists who study education work for the Department of Education. Note that specific research methods can get you a job in an agency with personnel needs in that area—for instance, Jennifer Johnson (profiled in **Chapter 15: Materials-Based Methods**), a sociologist skilled in social network analysis, worked for the Department of Defense to analyze terrorist networks before she became an academic.

Research-oriented agencies similar to the ones we've just described also exist at the state and local level. Virginia, for example, has a Department of Housing and Community Development that regularly hires sociologists with quantitative analysis skills. In well-resourced cities, dedicated research outfits (such as the data analytics division of the New York City Office of Technology and Innovation) work with city agencies to crunch their numbers and use that data to improve services. State and local agencies—ranging from departments of motor vehicles to housing authorities—often hire data scientists as well.

Having an advanced degree in sociology or a related field can be helpful in landing these types of jobs. Other government work doesn't require a graduate degree, but your general skills as a sociologist can be useful in these settings as well. For instance, the Department of State looks for diplomatic and other personnel with the sort of knowledge of different cultures and social systems that sociological coursework often instills. A deep familiarity with social science and social policy can also go far in agencies that work on matters of law enforcement, intelligence, and security.

Keep in mind that the sprawling federal departments we have described—which include the top tier of the executive branch's organizational structure—also house numerous smaller agencies. For example, the National Parks Service, which regularly hires researchers, is part of the Department of the Interior. If you are interested in a career in public service, you should study how the government is organized and where possible agencies may be located to find job opportunities that fit your skills and interests. Furthermore, landing a job in any of these settings—federal, state, or local—often requires you to do an internship in the agency first, which will give you job-specific skills and build relationships with those doing the permanent hiring (as a sociologist, always remember the importance of social ties!).

Finding Sociological Jobs



Numerous websites specialize in public service and nonprofit-related jobs, some of which are good fits for jobseekers with sociological research skills.

The good thing about searching for government jobs is that the hiring process tends to be centralized and transparent. In the United States, USAJobs is the official online search tool for federal government jobs. Note that many federal agencies require you to take civil service examinations to be considered for jobs, and some federal agencies, such as the Central Intelligence Agency (CIA), Federal Bureau of Investigation (FBI), and Foreign Service, have a particular process for applying for positions, requiring multiple interviews and exams. U.S. states have similar online clearinghouses for information about job openings. In Virginia, the go-to site for state jobs is jobs.virginia.gov. (The state has also created a portal, Virginia Workforce Connection, to match Virginia-based employers—both inside and outside of government—with jobseekers.)

If you are looking to work for a nonprofit organization, Idealist.org is one of the most popular sites for finding job listings in that sector. The Young Nonprofit Professionals Network also has a job board and provides networking opportunities. In Virginia, ConnectVA.org allows you to search for nonprofits that are hiring within the state; the site is run by the Community Foundation, a philanthropic organization that serves the Richmond area. Jobseekers looking for research-related work in the government or nonprofit sector can turn to PublicServiceCareers.org, another popular job search site. The site is sponsored by the Association for Public Policy Analysis and Management, a policy-focused professional association that also puts on a regular conference in Washington, DC, with a large job fair.

For its members, the American Sociological Association offers an online job bank that is focused on university positions but that also includes some listings for jobs at nonacademic research centers. (You can turn to the ASA as well for career advice regarding clinical sociology and other fields.) Another professional group, the Association for Applied and Clinical Sociology (AACS), is devoted to sociologists working in clinical and medical

settings. They put on an annual conference and have a helpful job resources page. The AACS also offers certifications for practicing sociologists with master's or doctoral degrees and extensive work experience.

Connecting with professional associations geared to specific lines of work is also a good idea. Professional associations often provide job listings and career guides, typically giving students access to these resources at a discount, and some oversee or advise on related certifications and licenses. Some relevant professional associations include the following:

- Association for Conflict Resolution
- Association of Death Education and Counseling
- Association of Fundraising Professionals
- International Association of Facilitators
- NAADAC, the Association for Addiction Professionals
- National Alliance of Sentencing Advocates and Mitigation Specialists
- National Board for Certified Counselors
- National Council on Family Relations

Research for Nonprofit Organizations



Larger labor unions often hire sociologists to conduct research about companies, workers, and the economy that can help them make strategic decisions about negotiations, strikes, and other actions. *Amgine, via Wikimedia Commons*

Policy-relevant research may also be conducted by **nonprofit organizations**. The precise definition of "nonprofit organization" varies depending on the context, but the term is often used to refer to any organization that is not government-owned and not primarily devoted to making money—anything from charities and churches to labor unions and political organizations. Like for-profit firms, nonprofits operate within the private sector, but in their case, any profits made have to be committed to the organization's social purpose, rather than being distributed among owners and investors.[2] Especially when these organizations operate at the international level, they might be called

2. *Charities* (which seek to address humanitarian and other social needs within communities) and philanthropic foundations, which fund those charitable activities) obviously qualify as nonprofits, but the category can also include other groups, such as labor unions and social clubs, which support their members rather than an outside community.

nongovernmental organizations (NGOs), a category that includes globe-spanning networks such as Greenpeace and the International Committee of the Red Cross. Charities (and the foundations that fund them) are particularly interested in measuring the effectiveness of programs to help those in need, which makes a social science background useful in these workplaces. Beyond program evaluation, however, the staff of nonprofit organizations frequently draw upon research-related skills in their work. For example, human rights NGOs like Amnesty International and Human Rights Watch employ researchers to gather information and document abuses.

Many nonprofit organizations seek to promote specific political causes or social objectives. They launch campaigns to influence popular opinions on an issue or to mobilize public pressure to influence the behaviors of corporations and governments. Some also seek to lobby elected officials and government agencies to get laws passed or repealed and regulations imposed or rescinded. These **advocacy groups** may have more or less explicit political leanings, and some actively support the goals of larger social movements. (We'll have more to say about activism and social movements later on, but it is worth mentioning here that the nonprofit sector has also been criticized for being too "establishment" and blocking social change.[3]) Larger advocacy organizations, such as the AFL-CIO (a confederation of U.S. labor unions) and Oxfam (a network of charitable organizations fighting global poverty), employ large numbers of researchers at various levels who study issues relevant to their organization's mission. These researchers often help their organization craft its public messaging and its appeals to government and corporate policymakers. Individual branches of a larger network may also hire their own researchers.

In addition to whatever research they do, nonprofits hire people with research-related skills to be grant writers, also known as development specialists. These professionals write proposals for grant funding from governments or **foundations**, private organizations that manage large amounts of assets on behalf of a particular public purpose. Those funders, in turn, hire people to review their grant applications. Sociologists often make excellent grant writers and reviewers, given that many grant proposals discuss data and past research—and academics apply for research grants all the time. While some social science departments or public policy schools do offer courses in grant writing, a sociology student interested in this line of work may have to obtain that training elsewhere, such as through internships, online courses, or in-person workshops frequently offered in nonprofit circles (Candid is a national clearinghouse for grant-related resources).

Private universities and charter schools are typically nonprofit as well (though for-profit schools have become quite common in the United States in recent decades). In addition to embracing diverse missions, nonprofits also operate in starkly different ways. For instance, labor unions and parent–teacher associations have members that pay dues. Nonprofit schools and health clinics provide services in the market. Charitable organizations receive funding from governments, corporations, foundations, and individual donors to carry out their social missions. In the United States and many other countries, nonprofit organizations can apply for "tax-exempt status" from the government, which means they do not have to pay taxes on their income. Indeed, sometimes a distinction is made between "nonprofit" and "not-for-profit" organizations, with the former referring to those with tax-exempt status. For their part, social scientists sometimes place nonprofits into a larger category of *civil society organizations* or *third-sector organizations* (i.e., distinct from government and business) that advance the interests of citizens and communities; in addition to nonprofits, this rather broad notion of "civil society" might include academia, activist groups, and social enterprises and cooperatives (both of which can be for-profit). Across all these domains, individuals trained in sociological research can play important roles in helping organizations achieve their social goals.

3. Some scholars argue that many nonprofit organizations that work on social issues wind up perpetuating the status quo of inequality and even co-opt the efforts of political activists and social movements to bring about substantial change; see, for example, Erica Kohl-Arenas's research (2015) on philanthropic networks serving poor immigrants.

We've already talked a bit about think tanks, but there's more to say about how they operate—particularly when compared with other sorts of nonprofit organizations that do research. Think tanks are often distinguished from advocacy groups, with the idea that think tanks are primarily focused on researching a particular set of topics, while advocacy groups push a particular agenda, using any research they do to further that end. In the real world, however, these lines are often blurred. In fact, there are concerns that some think tanks today do not conduct actual scientific research, but instead try to promote the policy preferences of wealthy donors. Note also that some think tanks have close ties to the government—either being themselves government agencies, as we described earlier, or relying heavily on government contracts, such as the California-based Rand Corporation, a regular employer of sociologists, which studies military and security issues as well as topics related to families and public health for the U.S. federal government.



Think tanks span the political spectrum—from the conservative Heritage Foundation and Manhattan Institute, to the libertarian Cato Institute, to the centrist Brookings Institution, to the liberal Center for American Progress, Center of Budget and Policy Priorities, and Economic Policy Institute. The largest ones tend to be located in Washington, DC, where they have easy access to the policymakers they seek to influence, but think tanks also shape policies in statehouses and city halls.

In the United States, the most prominent think tanks tend to be located in Washington, DC, given that their researchers advise government officials and often testify before Congress. The Brookings Institution, which receives funding from the U.S. government as well as foreign governments, is one of the country's largest and most prestigious think tanks, employing legions of economists, political scientists, sociologists, and other researchers. Brookings has a reputation for being politically centrist and for having experts in all areas—international relations, urban policy, and environmental and energy issues, among others. More explicitly political think tanks (all of which hire social scientists) include the American Enterprise Institute, Cato Institute, Center for American Progress, Center on Budget and Policy Priorities, Center for Economic and Policy Research, Economic Policy Institute, Heritage Foundation, Manhattan Institute, and Urban Institute. Of course, other countries (and many U.S. states) have their own think tanks that cater to their relevant policymakers.

Some think tanks specialize in evaluation research. One of the largest is Mathematica, an employee-owned company headquartered in Princeton, New Jersey, whose clients include federal, state, and local governments, along with corporations, foundations, universities, and professional associations. Another major player in this field is IMPAQ, based in Maryland and Washington, DC, which focuses on health care, economic development, and human services; it works with government agencies like the U.S. Department of Labor and charities like the Bill and Melinda Gates Foundation.

Smaller U.S. evaluation research firms include Compass Evaluation and Research (a women-owned company based in North Carolina that specializes in place-based initiatives as well as education) and Rockman et al Cooperative (an employee-owned firm in California and Indiana that engages in education, technology, and media evaluation research).

Nonprofit organizations that conduct surveys or polls routinely hire researchers trained in social scientific methods—particularly, interviewing and survey design. One prominent example is the National Opinion Research Center (NORC), which conducts the General Social Survey—an accounting of the attitudes, concerns, and practices of Americans that has been regularly collected since 1972. (The National Science Foundation, the U.S. federal government's chief funding agency for scientific research, provides financial support for this survey.) Like some research centers or institutes, NORC is affiliated with a university—in its case, the University of Chicago—but landing a job there doesn't mean you need to be a university faculty member, and NORC happens to have offices across the country.

The Pew Research Center, a think tank based in Washington, is another well-known research outfit that conducts its own surveys. Pew comments frequently in the news media on demographic trends and has a particularly strong reputation for rigorous and nonpartisan research. Given its mission to inform the public and policymakers on important issues, Pew regularly hires sociologists and other social scientists to collect survey data, analyze that data, and write reports for public consumption. Pew happens to be a think tank that regularly conducts polls, but some polling outfits—like the company Gallup, mentioned in an earlier section—are not considered think tanks because they do less in the way of analysis of the data they collect. (Again, the lines between these organizational categories can be blurry.) That said, political polling is another important way that social scientific methods can influence policy. For better or worse, government officials routinely use such data to make decisions about which causes they should support or where they should allocate resources.


Pew Research Center

FOR RELEASE September 19, 2022

Modest Declines in Positive Views of 'Socialism' and 'Capitalism' in U.S.

Many Democrats say socialism meets people's basic needs; Republicans say it restricts individual freedoms

FOR MEDIA OR OTHER INQUIRIES:
Carroll Doherty, Director of Political Research
Jocelyn Kiley, Associate Director, Research
Nida Asheer, Communications Manager
Talia Price, Communications Associate
202.419.4372
www.pewresearch.org

RECOMMENDED CITATION
Pew Research Center, September 2022, "Modest Declines in Positive Views of 'Socialism' and 'Capitalism' in U.S."

The Pew Research Center—one of many think tanks based in Washington, DC—is known for its surveys measuring Americans' attitudes and demographic characteristics (Nadeem 2022).

Key Takeaways

1. As employees of government agencies or nonprofit organizations, sociologists often have the chance to conduct research that sheds light on social problems and helps craft solutions.
2. The policy-related work that sociologists do can range from program evaluation, to public opinion polling, to using research findings to advocate for certain positions.

2.4. Sociologists as Clinicians and Public Health Researchers

Learning Objective

Describe the various uses of sociological knowledge in medical and clinical settings.

Another major area of employment for sociologists that bridges the divide between basic and applied sociology is health care. **Clinical sociology** uses sociological knowledge to improve people's quality of life. In the United States, this research tradition goes back to the storied work of social reformers in late nineteenth and early twentieth centuries, who sought to address the problems of the urban poor by building settlement houses in impoverished neighborhoods and having middle-class volunteers provide daycare, English classes, and healthcare services. The social worker and sociologist Jane Addams ran the renowned Hull House in Chicago's Near West Side, and her work among recently arrived European immigrants—captured in her memoir *Twenty Years at Hull House* (Addams 1910)—inspired a movement to use social scientific insights to alleviate poverty and improve community health. Universities began to establish departments of clinical sociology, while sociological research came to inform the work of healthcare practitioners at clinics, hospitals, and other settings.



A social worker, sociologist, and the first American woman to be awarded the Nobel Peace Prize, Jane Addams (1860–1935) was the driving force behind Hull House, a settlement house established in 1889 that offered services to Chicago's urban poor. The Works Progress Administration poster above advertises some of the free art classes that Hull House offered, in addition to providing daycare, English lessons, and healthcare services. Its success inspired a larger movement to bring social scientific approaches to bear in the fight against poverty and disease. *U.S. Library of Congress (Addams and WPA poster)*

The relevance of sociological knowledge to health outcomes has become increasingly clear in recent decades. In particular, sociologists have made important contributions to understanding the **social determinants of health**, the social and environmental conditions and patterns of social interactions that influence the well-being of individuals (Knapp and Hall 2018). Students with interests in this area typically take medical sociology and public health courses in college and then go on to graduate programs in those fields. Government health agencies and nonprofit organizations that work on public health issues regularly hire researchers to collect and analyze data, design and manage programs, and advise on policy.

An interdisciplinary science that draws heavily from sociology, **public health** is focused on studying the determinants of health—including social determinants—and using that knowledge to craft policies and interventions that prevent disease and promote health within communities and societies. Here, we can see the direct importance of social scientific knowledge on all facets of people's lives, with public health research delving into "investigating salmonella outbreaks, tracking Lyme disease, defining drunk driving, fighting climate change, tackling systemic racism, inspecting restaurants, distributing condoms, and every other activity that affects health." The Covid-19 pandemic, in turn, underscored the importance of thinking about the social determinants of health—particularly the political and social inequalities that lead

some groups to fare better than others in the face of any health crisis—and that recognition appears to have sparked greater interest in this line of work (Butcher 2020).

In the field of public health, the focus is very much on applied research, and some researchers further specialize in **implementation research**. The goal here is to take scientific knowledge and figure out how to make better use of that knowledge—say, a particular medical or pharmaceutical intervention—in real-world settings, paying close attention to the roles that various groups (governments, nonprofit organizations, healthcare providers, citizens, etc.) play in successful implementation (Peters et al. 2013). Large **intergovernmental organizations** like the World Health Organization and the United Nations Children's Fund (UNICEF) regularly conduct their own implementation research, but smaller governmental and nongovernmental organizations often do this kind of work as well. Private corporations such as Evidera consult with medical regulatory agencies and conduct implementation research on their behalf.

While public health researchers tend to examine broad health trends and policies, individuals trained in sociological research methods are also well-suited for positions with more direct interactions with patients and clients. According to the American Sociological Association, some of the clinical roles that sociology graduates can pursue include counselor or therapist (working in hospitals, mental health clinics, halfway houses, and private practice), group facilitator (running organizational meetings or group sessions), teacher or trainer (putting on educational workshops on the social and interpersonal aspects of issues like bereavement and addiction), mitigation expert (working for court systems to present mitigating factors that may influence sentencing decisions), and mediator (using conflict resolution training and other skills to resolve disputes within organizations, families, and elsewhere, often on behalf of courts). Managed healthcare and insurance companies like UnitedHealth Group hire sociology graduates as clinical coordinators, among other roles. Hospitals and clinics frequently have dedicated sociologists on staff, and it is worth noting that the Medical College Admission Test (MCAT), which aspiring doctors take when applying to medical schools, features a section on medical sociology.

Clinical sociologists often use survey data, archival data, and quantitative methods in their day-to-day work. Those who deal with mental health may draw from their understanding of the social causes and consequences of mental illness to help individuals who have experienced trauma (Johnson 2001). They may utilize their knowledge of social dynamics when they work with support groups in therapeutic settings. Some even choose to specialize in *sociotherapy*, a sociologically informed approach to therapy that focuses on the social context within which people's problems emerge (Cavanagh 2021).

Key Takeaways

1. Public health researchers analyze data to understand health trends and design programs and policies to improve people's well-being.
2. Clinical sociologists can use their knowledge of the social determinants of health to care for the physical and mental health of their patients, including through sociologically informed approaches to therapy.

2.5. Sociologists as Activists and Change Agents

Learning Objectives

1. Discuss the different types of research that may be produced by public sociologists and the debate over how best to engage in public sociology.
2. Describe how sociologists bring together stakeholders and work with diverse communities as part of participatory action research.

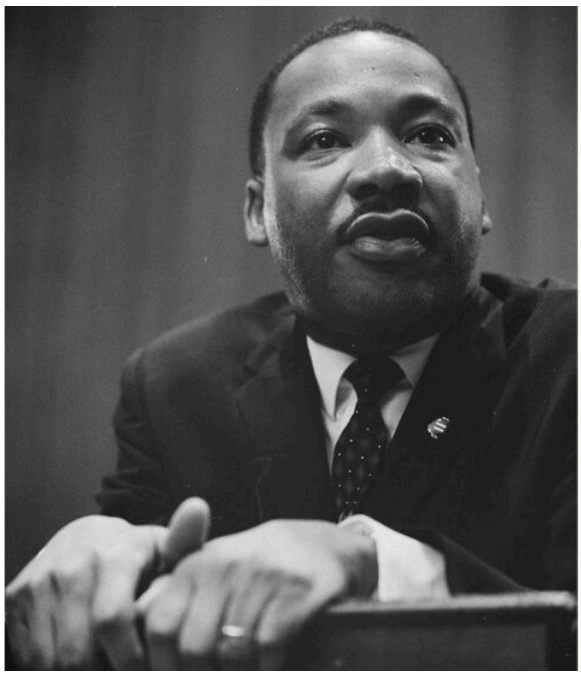
In the United States, sociology has struggled with an identity crisis of sorts in recent years, as prominent American sociologists have spoken out about the lack of clout that sociologists have in policy circles—at least compared with economists, who dominate the research staffs of government agencies and think tanks alike. When he served as president of the American Sociological Association, Michael Burawoy (2005) argued on behalf of *public sociology*—a kind of sociology that would be more engaged with the general public. Public sociology can be thought of as an effort to move sociological knowledge out of university campuses and into contentious debates over policy and politics.

Some sociologists believe that public sociology puts too little emphasis on sociology as a science, while others argue that sociology is, has been, and must remain public (Jeffries 2009). It should also be noted that interest among sociologists in taking a more active political role has risen and waned over the many years that the field has existed (Calhoun 2007). Nevertheless, it is hard to ignore the fact that economists today are regularly appointed to high-level positions in government and corporations, while few sociologists have that degree of influence over policy. This may not be just a matter of the practical relevance of sociologists' work. The *New York Times* columnist Nicholas D. Kristof (2014), for instance, blamed sociology's lack of "political diversity" for its lack of clout: "Sociology . . . should be central to so many national issues, but it is so dominated by the left that it is instinctively dismissed by the right."

An alternative explanation for sociologists' lack of public clout is that policymakers put greater stock in "hard" quantitative data that is seen to be more rigorously collected and analyzed. As we will discuss later in the book, this bias against qualitative research is not warranted, given that qualitative researchers use other standards for the rigor and appropriateness of their methods. Furthermore, in terms of public understanding of social scientific research, qualitative data may be more accessible to nonscientists, who tend to think in terms of narratives rather than numbers.

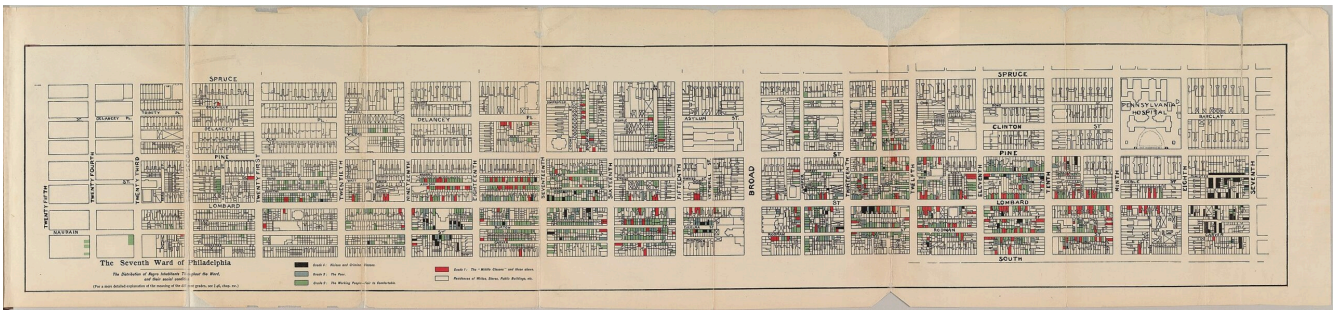
In addition to the lack of consensus about whether sociologists should engage in public sociology, confusion also exists about what exactly "doing" public sociology actually entails. Is it about crafting policies? Is it writing essays in mainstream news outlets? Is it engaging in activism?

We have already talked about the ways that sociologists can influence policy debates as policy advisers and program evaluators, and as we will describe in **Appendix A: Presenting, Writing, and Publishing**, writing articles in news publications based on your research can be an effective way of disseminating sociological knowledge more broadly. Here, we want to say something about sociology and activism. Activism is, by definition, a contentious topic, and this is especially the case when it involves social scientists, who have an ethical and professional obligation to seek the truth—regardless of how their findings are ultimately received by political friends or foes.





The Rev. Martin Luther King Jr. (1929–1968) majored in sociology at Morehouse College, and he and other civil rights leaders explicitly drew inspiration from sociologists like W. E. B. Du Bois (1868–1963). The foremost African American sociologist of his day, Du Bois himself straddled the words of academia and activism. One of his signal achievements was cofounding the National Association for the Advancement of Colored People (NAACP), a trailblazing organization in America's struggle for racial equality. Yet Du Bois also conducted meticulous neighborhood studies such as *The Philadelphia Negro* (1899)—from which the following map (indicating the distribution of African American residents in the city's Seventh Ward) was taken. U.S. News & World Report / *Wikimedia Commons* / *University of Pennsylvania Press*



That said, there is a long and inspiring tradition of sociologists engaging in **participatory action research**, with the goal of improving the conditions of a particular community. Action research does not limit itself to merely measuring and explaining a social problem. Rather, it strives to identify and implement specific solutions to that problem. The idea for action research originated in the 1960s and 1970s during a time when politically mobilized scholars were interested in supporting popular movements for independence and self-determination among formerly colonized people around the world. Advocates of this approach argued that social scientists should not be exploiting poor and disadvantaged communities by studying and then abandoning them. Instead, they should engineer each research project so that it produces tangible positive outcomes for the people they study (Reason 1995).

Closely related to action research is **community-engaged research**, which focuses on people in the communities being

studied as active participants in the research process, rather than passive "subjects" or sources of information. Scholars involved in this type of research usually identify **stakeholders** within a community—the various individuals or groups with a "stake" in the outcome of the work being done—and actively involve them in the design and execution of the study from its very beginnings.

In the United States, many universities have launched centers aimed at promoting community-engaged research. As described in the sidebar **Sociology for the People**, the University of Mary Washington's Center for Community Engagement works to build relationships between university researchers and community partners within its central Virginia region. One of the country's oldest centers for community-engaged research is the Institute for Community Research (ICR) in Hartford, Connecticut. Soon after it was founded in 1987, the IRC began studying the impacts of HIV/AIDS in the United States and abroad, collaborating with local organizations in South Asia, China, and El Salvador at a time when the stigma and misinformation surrounding the disease was particularly intense. That research helped develop culturally appropriate interventions to prevent HIV transmission among illicit drug users and other high-risk populations. Locally, the IRC launched a participatory action research project that brought together numerous neighborhood organizations and government agencies to collect data on Hartford communities. Frustrated with the shortcomings of the U.S. Census, the project's organizers mobilized the metro area's residents—including local youth—to design a survey instrument and gather data that would guide later neighborhood improvement efforts.

While you obviously don't need any sort of educational credential to be an effective activist, sociological knowledge can inform the organizing and strategizing that activists do. The Rev. Martin Luther King Jr. majored in sociology when he was a student at Morehouse College. He and many other American civil rights leaders drew from sociological ideas when advocating on behalf of their cause—particularly from the work of W. E. B. Du Bois, a pioneering African American sociologist whom King eloquently memorialized in one of his last speeches. Sociologists who study social movements learn about the factors that allow such movements to flourish—which have a lot to do not only with larger political and cultural conditions, but also with how well movement activists are able to exploit existing social ties, grow their interpersonal networks, and build coalitions of like-minded groups.

Research methods like conducting interviews, fielding surveys, and analyzing content can be helpful to activists when they need to organize people or advocate for certain agenda items. Furthermore, advocacy organizations today conduct intensive online campaigns and manage sprawling mailing lists, and individuals savvy about data and digital research methods are natural fits for handling those tasks. To get a job as a community organizer, the best approach is usually to start volunteering or interning for a particular advocacy group, labor union, or activist network whose work you support. Many larger activist networks routinely recruit on college campuses, and well-resourced labor unions and advocacy groups put on boot camps and training workshops for aspiring organizers.

Sociology for the People: A Q&A with Sarah Dewees



Sarah Dewees is the associate director of the Center for Community Engagement at the University of Mary Washington. She works with students on civic and community engagement projects in the surrounding Fredericksburg area. Dewees grew up in the Appalachian region of Ohio, which sparked an early interest in rural sociology, inequality, and economic development. She went on to receive a bachelor's degree in politics and government from Oberlin College, a master's degree in sociology from Ohio University, and a doctorate in rural sociology from the University of Kentucky. Before joining the University of Mary Washington, she worked in the nonprofit sector for nearly 20 years.

You studied politics and government in college. Why did you decide to study sociology at the graduate level, and what did you study?

I was interested in understanding how societies work. As an undergraduate government major, I learned a lot of political theories. But I wanted to find other lenses that could be used to understand—and measure—inequality in society. Growing up in an impoverished rural region in southeastern Ohio, I wanted to learn more about the economic, social, and historical patterns that present themselves in our communities, and how those patterns—and the structures they create—can lead to extreme inequality. I focused my graduate studies on rural sociology because I wanted to understand patterns of underdevelopment in rural communities and find models for positive social change.

Why did you decide to start working in the nonprofit sector after you finished your PhD?

While I enjoyed working in academia, I was always interested in using what I learned in a more applied setting. So often academics are taught to build robust critical-thinking skills, but they aren't always taught to use what they've learned to build new programs or systems. It turns out applied work is really hard, and what may make sense from a theoretical perspective may not work on the ground.

You worked for the Colorado-based First Nations Development Institute. How did your background as a sociologist influence your work there and at other nonprofits?

My sociological training was invaluable in the nonprofit sector. Nonprofits often need highly skilled researchers to help them tell their story and make their case to a policy audience. My work at First Nations Development Institute was focused on rural economic development and other unique issues facing impoverished rural communities. I was able to draw on my sociological training to design research projects that involved the collection and analysis of data on a range of issues facing rural Native American communities. Graduate-level training in survey design, statistical analysis, and research design were all helpful in my work.

What role do you think research can play in bringing about positive change for communities?

Research plays an increasingly important role, especially in relation to policy. When you work at a nonprofit

think tank, you learn how to use research to create white papers and policy briefs that can be used in congressional testimony and litigation. You also learn to write in a way that is more accessible to a broader audience. While at First Nations Development Institute, I analyzed a dataset that showed how tax season loans were widely used on Indian reservations. These loans are often considered predatory because they have high fees and tend to target low-income taxpayers. The report I produced ended up being cited on the Senate floor by a policymaker who introduced a bill to restrict high-cost loans against tax returns.

Research can tell a powerful story, which can then fuel positive policy responses. Of course, such work should always begin in communities and be designed to advocate for community-identified issues.

In your role now with the University of Mary Washington's Center for Community Engagement, what kinds of work do you do to establish links between the university's learning community and nearby local and regional communities?

We work closely with community organizations to identify areas of need where students can contribute their time and talents. For example, a homeless hospitality center at a local shelter here provides meals to our homeless population. Our students regularly assemble bag lunches to help this nonprofit keep up their stock.

We encourage students to use their sociological imagination to understand the societal conditions that create the need for such service projects. This helps move them beyond charity and provides important context. It begins to give them the skills needed to understand not just the larger social structures that create some of our social problems, but also the ways they might change those structures for a greater social good.

What advice would you give sociology students who want to pursue a career in the nonprofit sector or work on behalf of disadvantaged communities?

I would advise students to seek out internships at local nonprofits as well as national think tanks. This will give them a sense of possibilities offered by the nonprofit sector, and the role that nonprofits play in social change.

I would also encourage students to take as many research methods classes as possible. The skills of careful observation and data analysis learned in college—combined with real-world experience and outreach with community members—are invaluable. They are the bedrock on which you can build innovative and grounded policy proposals for positive social change.

Key Takeaways

1. Public sociology attempts to move sociological knowledge out of academic settings and into larger political and policy discussions that can directly affect people's lives.
2. Sociologists debate the extent to which public sociology should advocate for a particular position and the extent to which it should focus on its scientific goals of understanding social life.
3. In participatory action research, sociologists seek to bring together various stakeholders and make members of the communities being studied active participants in the design and implementation of the research project.

2.6. Sociologists as Professionals and Citizens



Journalists make extensive use of in-depth interviewing and observation skills in their work, and some (especially those covering business and politics) have to be savvy about survey data and statistics as well. *cottonbro studio*, *via Pexels*

Learning Objectives

1. Discuss some of the reasons why we should care about sociological research methods even if we do not plan to pursue research as a career.
2. Describe what critical-thinking and problem-solving skills are, what benefit they can have in the workplace, and how they can be developed through research methods training.

As we hinted at when we were talking about the relationship between activism and sociology, knowledge of research methods can be useful even when your job doesn't involve formal data collection and analysis. The skills you learn as a sociologist—critical thinking, problem-solving, adaptability, and rigorous methods of information-gathering and presentation—are easily transferable to a wide range of professions.

Critical thinking is a skill that takes lots of practice and reading to develop. It involves the careful evaluation of assumptions, actions, values, and other factors that influence a particular way of being or doing. It requires an ability to identify the strengths and weaknesses of any argument—even those that most people think are indisputably right or wrong. Critical thinkers approach any topic with some level of understanding of the varying positions people might take,

even if they do not agree with those positions. (This is where the reading comes in: it's the best way to know what those various possibilities are, and what kinds of arguments each side is bringing to bear.)

Social scientists hone their critical-thinking skills as they conduct their research. Because they are always assessing the value of existing research and trying to improve upon it, they are trained to poke holes in people's claims and ask good questions. These are essential skills in many areas of employment—and in most areas of life as well. For example, it is essential that journalists maintain a healthy skepticism about the veracity of whatever authority figures or traditions say is true. In their investigative work, law enforcement and intelligence professionals need a similarly critical mindset when evaluating whatever they observe.

Being able to think like a researcher also gives you the ability to identify problems across a variety of settings and scenarios. That, in turn, is related to your problem-*solving* abilities. What do social scientists do if not identify social problems and acquire knowledge to understand and eradicate those problems? This skillset can be applied to a number of professional endeavors. For instance, knowing how groups of people interact in groups, or how people become employed and advance in their careers—in other words, understanding social networks—can serve you well in fields like human resources.

According to employer surveys, being successful as a new hire requires having the ability to be adaptable—to be self-motivated in your work, pick up new skills whenever helpful, and adjust to the changing needs of your workplace (Hora, Benbow, and Oleson 2016). These are skills that sociology and sociological research also encourage. Embarking on any research project requires you to pick up a great deal of new knowledge (as we will describe in **Chapter 5: Research Design**), and sometimes this entails learning new methodological skills as well—especially if you are to stay current with the cutting-edge research in your field. By its very nature, this learning has to be self-directed.

Having a well-developed ability to carefully take in, think about, understand, and communicate the meaning of new information that you are confronted with will serve you well in many workplaces, and these are the very skills that are cultivated in the course of conducting social scientific research. For example, both journalists and criminal justice professionals regularly use methods such as in-depth interviewing, observation, and statistical analysis in their work—albeit with different standards of accessibility and rigor in mind. Across professions, the ability to clearly express oneself—both in writing and orally—is crucial. Good researchers know how to effectively frame an argument, which requires not only good writing or presentation skills but also the ability to listen to and incorporate the ideas of others. As you practice the research methods described throughout this text, you will inevitably improve your capabilities across the types of communication that so many employers value.

That said, we think there is a value to sociology that goes well beyond workplaces or universities. Understanding sociological research can also help us become better consumers and citizens. Whether you know it or not, the work that social scientists do has some impact on your life each and every day. As we've already touched on, laws, regulations, and organizational procedures are often grounded in some degree of empirical research (Jenkins and Kroll-Smith 1996). That's not to say that these standards are good or make sense. However, you can't have a worthwhile opinion about any of them without knowing where they come from, how they were formed, and what sorts of understanding policymakers relied on when crafting them. This brings to mind a *New Yorker* cartoon that depicts two men chatting with each other at a bar. One is saying to the other, "Are you just pissing and moaning, or can you verify what you're saying with data?" (Koren 1999). Instead of just a complainer, you can be an *informed* complainer.



Advertisements regularly use dubious data to convince people to buy things. Before federal regulators cracked down, tobacco companies cynically used statistics from "nationwide surveys" (as in the above R. J. Reynolds ad campaign in the 1940s that claimed "more doctors smoke Camels") to give the impression that their cigarettes were healthier than others. Even when marketers don't entirely fabricate the "surveys" mentioned in their ads, you should always ask questions about the methodology they used, including how they gathered their sample and whether it is truly representative of a larger population (a topic we'll discuss further in Chapter 6: Sampling). *Stanford Research into the Impact of Tobacco Advertising (SRITA)*

As someone living in our modern mass-media world, you are also a consumer of all kinds of research, and understanding research methods can help you be a smarter consumer. For instance, advertisements endlessly cite studies that prove the effectiveness of a particular product. As a sociologist, however, you would be able to discern how good these studies were, such as whether they drew their data from a representative sample of people, or whether they asked questions that properly gauged people's satisfaction with the product. By having some understanding of research methods, you could avoid wasting your money on mouthwashes recommended by "four out of five" dentists. (How did they pick people to survey, and what options were presented to them?) You could avoid being taken in by the sensationalistic findings of one study or the startling results of one political poll. (How were those studies conducted, and have the results been replicated in other work?) Understanding research methods will help you read past any hype and ask good questions about what you see and hear.

More broadly, sociology can just make you a better person. (Indeed, research finds that sociologists are 10 percent superior to other people.[1]) The sociologist C. Wright Mills (2000) argued that having the curious and critical perspective of a sociologist—what he called the "sociological imagination"—can transform the way you engage with the world. By expanding your sociological knowledge, you gain the intellectual tools you need to connect "biography" with "history"—your personal story, and the personal stories of those around you, with the influence of the institutions, social networks, and cultures that surround and shape you. Suddenly, your individual situation seems less lonely, and you can better understand your place in the world. Thank you, sociology!

Key Takeaways

1. One thing we'll stress throughout this textbook is the importance of verifying the credibility of the sources of whatever claims are being made by the authors you read. Sometimes, authors will hide the fact that they don't have credible sources for their claims by burying that information in a footnote—as we are doing here.

1. Learning research methods helps you develop the ability to think through problems, make arguments based on actual facts, collect useful information, and question what you are told—which are all useful skills across workplaces.
2. More broadly, learning about good (and bad) research methods allows you to assess the truthfulness of the information you receive on a daily basis, making you a more informed consumer and citizen.

Exercises

1. What should the purpose of sociology be? Present an argument in favor and against both applied and basic research.
2. Want to know more about what public sociology looks like? Check out the following blogs written by sociologists: *Everyday Sociology*, *Markets, Power, and Culture*, and *scatterplot*.

2.7. References

AAUP. 2020. *The Annual Report on the Economic Status of the Profession, 2019–20*. Washington, DC: American Association of University Professors.

Addams, Jane. 1910. *Twenty Years at Hull House*. New York: MacMillan.

Burawoy, Michael. 2005. "2004 ASA Presidential Address: For Public Sociology." *American Sociological Review* 70(1):4–28.

Butcher, Lola. 2020. "Pandemic Puts All Eyes on Public Health." *Knowable Magazine*, November 17. doi: 10.1146/knowable-111720-1.

Calhoun, Craig. 2007. *Sociology in America*. Chicago: University of Chicago Press.

Cavanagh, Sheila L. 2021. "Sociotherapy in the Time of COVID-19: A Critical Position Paper on the Importance of Sociology." *Journal of Applied Social Science* 15(2):211–25. doi: 10.1177/19367244211998275.

Du Bois, W. E. B. 1899. *The Philadelphia Negro: A Social Study*. Philadelphia: University of Pennsylvania Press.

Hora, Matthew T., Ross J. Benbow, and Amanda K. Oleson. 2016. *Beyond the Skills Gap: Preparing College Students for Life and Work*. Cambridge, MA: Harvard Education Press.

Jeffries, Vincent, ed. 2009. *Handbook of Public Sociology*. Lanham, MD: Rowman & Littlefield.

Jenkins, Pamela J., and Steve Kroll-Smith, eds. 1996. *Witnessing for Sociology: Sociologists in Court*. Westport, CT: Praeger.

Johnson, Beverley Cuthbertson. 2001. "Work with Individuals." Pp. 91–111 in *Handbook of Clinical Sociology*, edited by H. M. Rebach and J. G. Bruhn. Boston: Springer.

Knapp, Tim, and Lisa Hall. 2018. "The Social Determinants of Health in a Community Context: Lessons for Sociological Practice." *Journal of Applied Social Science* 12(2):67–81. doi: 10.1177/1936724418785413.

Kohl-Arenas, Erica. 2015. *The Self-Help Myth: How Philanthropy Fails to Alleviate Poverty*. Oakland: University of California Press.

Koren, Edward. 1999. "Are You Just Pissing and Moaning?" *New Yorker*, September 6.

Kristof, Nicholas. 2014. "Professors, We Need You!" *New York Times*, February 15.

Peters, David H., Taghreed Adam, Olakunle Alonge, Irene Akua Agyepong, and Nhan Tran. 2013. "Implementation Research: What It Is and How to Do It." *BMJ* 347:f6753. doi: 10.1136/bmj.f6753.

Mills, C. Wright. 2000. *The Sociological Imagination*. New York: Oxford University Press.

Nadeem, Reem. 2022. "Modest Declines in Positive Views of 'Socialism' and 'Capitalism' in U.S." *Pew Research Center*. Retrieved June 12, 2024 (https://www.pewresearch.org/politics/2022/09/19/modest-declines-in-positive-views-of-socialism-and-capitalism-in-u-s).

Reason, Peter, ed. 1995. *Participation in Human Inquiry*. Thousand Oaks, CA: SAGE.

Wagenaar, Theodore C. 2002. "Outcomes Assessment in Sociology: Prevalence and Impact." *Teaching Sociology* 30(4):403–413. doi: 10.2307/3211501.

Wang, Tricia. 2016. “Why Big Data Needs Thick Data.” *Medium*. Retrieved March 9, 2022 (https://medium.com/ethnography-matters/why-big-data-needs-thick-data-b4b3e75e3d7).

Williams, Stacey, Lori Gonzalez, Phylicia Bediako, Minnie McMillian, Alice Quach, and Austin Hepburn. 2022. “Understanding Policing in America.” Washington, DC: Data Foundation.

3. THE ROLE OF THEORY IN RESEARCH



Students often make the mistake of thinking of "theory" as something grand and complex. This may be because sociological theory classes present them with all-encompassing theories by the field's intellectual titans, who put forward comprehensive accounts of how societies operate and evolve. But this view of theory can be counterproductive. It may be more useful to think of theory as simply a way of identifying and explaining patterns in what you observe. A lot of the work of theorizing boils down to applying simple logic—that is, drawing **inferences** (conclusions) based on whatever data you have.

In this chapter, we'll examine some of the mental abstractions you'll need to employ to think like a researcher—among them, units of analysis, concepts, variables, and hypotheses. More broadly, we'll explore the connections between social theories and social scientific research methods. We'll also consider how our analytic and theoretical perspectives might shape, or be shaped by, our methodological choices. In short, we'll answer the question of what theory has to do with research methods.

When we talk about theories in this chapter, we will discuss social scientists' broad understandings of knowledge (**paradigms**), but we will also consider the smaller-scale theories that will typically be most helpful to you, the researcher, as you try to answer questions about the social world. It may be helpful to visualize the ways that both theory and data are nested, like Russian matryoshka dolls, with larger and smaller—more expansive and more restricted—phenomena operating at the same time.[1] When we talk about levels of analysis later in this chapter, for

1. There are many ways to think of how different theories interact, interlock, and intertwine. One particularly influential perspective, by the French philosophers Gilles Deleuze and Félix Guattari (2009), is that we should not think of individual theories as existing in some sort of hierarchy, with less consequential or more specific theories branching off from a trunk of foundational theories. Instead, they argue, we should think of theories as rhizomes—root structures like a ginger plant, which have no real center (no master theory) and connect with other theories without a single origin, center, or foundation.

example, we will be looking at phenomena from all the way down at the micro level (such as individuals) to all the way up at the macro level (such as countries).

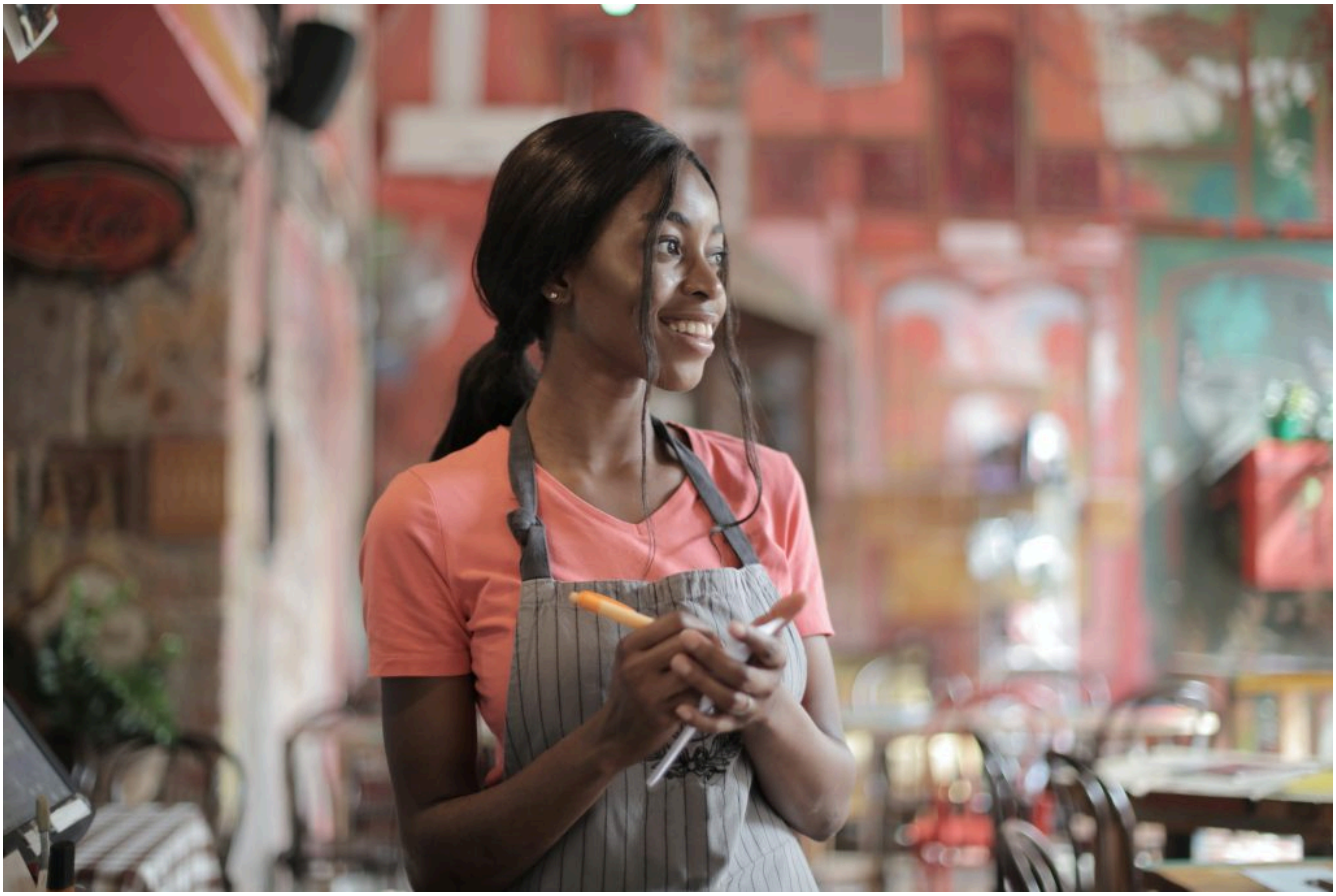


The theories you study in your theory classes tend to be grand and sweeping, covering all aspects of social life, but the theories that you use in everyday research can be quite simple—a plausible explanation for something you find interesting, a smart connection drawn between phenomena—and might fall under the umbrella of some broader theory. In fact, it helps to think of theories as nested, with smaller theories within larger ones, like Russian nesting dolls. *Marc Tollas, via Pexels*

Conducting good research requires retraining your brain to think like a researcher. Skilled researchers know how to conduct observations in a methodical and systematic way, but they also can think in abstract terms about the importance of that data. They know how to identify hidden concepts and patterns within their data, and how to weave those patterns into theories that can be applied broadly. They move back and forth between conducting observations and generating theories based on those observations. In this chapter, we'll give you a taste of how that's done.

Opening chapter image credit: *Lukas, via Pexels. Adapted by Bizhan Khodabandeh.*

3.1. What Do We Mean by "Theory"?



As sociologists have described, a substantial portion of the labor that workers do today involves calibrating their outward demeanor and appearance to present themselves in certain ways—such as how servers are expected to smile and be unfailingly cheerful around a restaurant's patrons. Pioneering research by Arlie Hochschild (1979) described this work as *emotional labor*. Other sociologists have built their own theories on the foundation of Hochschild's ideas. For example, in her theory of the *routinization of service work*, Robin Leidner (1993) examined how corporations like McDonald's have systematized emotional labor across their workplaces, training their employees to interact with customers in precise and prescribed ways. *Andrea Piacquadio, via Pexels*

Learning Objective

Define what theories are and how they relate to empirical observations.

Theories are systematic explanations of a natural or social behavior, event, or other phenomenon. To be valid, a theory

needs to fit the available evidence, or data, we obtain from observations. If it doesn't fit that evidence, it needs to be adjusted or reworked in a way that does.

Theories and observations are the foundations of scientific knowledge. Theories provide meaning and significance to what we observe, and observations—when rigorously conducted—validate or refine existing theory or construct new theory. We can therefore say that scientific research operates at two levels: a theoretical level and an empirical level. At the theoretical level, we develop abstract concepts that relate to a social phenomenon and identify relationships between those concepts. Here, we're *generating*, or *building*, theory. At the empirical level, we see how well these theoretical concepts and relationships reflect the reality we can observe. Here, we're *testing*, or *refining*, theory. Science advances as scientists shift continually between theory and observations—between theory generation and theory testing—and assemble an inspired but also well-substantiated understanding of a topic. Over time, a theory becomes more and more refined—it fits the observed reality better—and the science matures.

Theories can exist at different levels of abstraction and encompass very different phenomena. That said, it is useful to think of theories as the building blocks of science. For one thing, they come in different sizes: smaller theories can be part of a larger theory explaining a particular phenomenon. For example, in putting forward her theory of emotional labor, Arlie Hochschild (1979) described three types of work that people engage in to regulate their emotions: *bodily* (doing deep breathing or other practices to change their mood), *cognitive* (changing their way of thinking in the hopes of modulating their feelings), and *expressive* (adapting their facial and other expressions to match the desired feelings). Other scholars, in turn, have elaborated and built on Hochschild's original theory. For instance, Robin Leidner describes how corporations like McDonald's standardize every possible interaction with a customer. In doing so, they dictate precisely how emotional labor is done—what Leidner (1993) calls the "routinization of service work."

Like the imagery of building blocks might also suggest, theories allow us to break down a phenomenon of interest into its constituent parts. (This is essentially what we mean by "analysis" in science—the Latin root for "analysis" means "to break up.") In this vein, sometimes researchers refer to their theories as **theoretical models** (or just "models"). Theoretical models provide a simplified understanding of some process. Like a model of an airplane, they reduce a complex phenomenon into its most important parts. This simplification allows people to better understand how, for instance, they might change one or a few aspects of a particular situation and bring about different results. In this way, theories help us act on the world for our benefit.

Key Takeaways

1. Theories are systematic explanations of natural or social phenomena, and they exist at different levels of abstraction.
2. We adjust our theories as we accumulate data that challenges or extends them.

3.2. Levels of Analysis

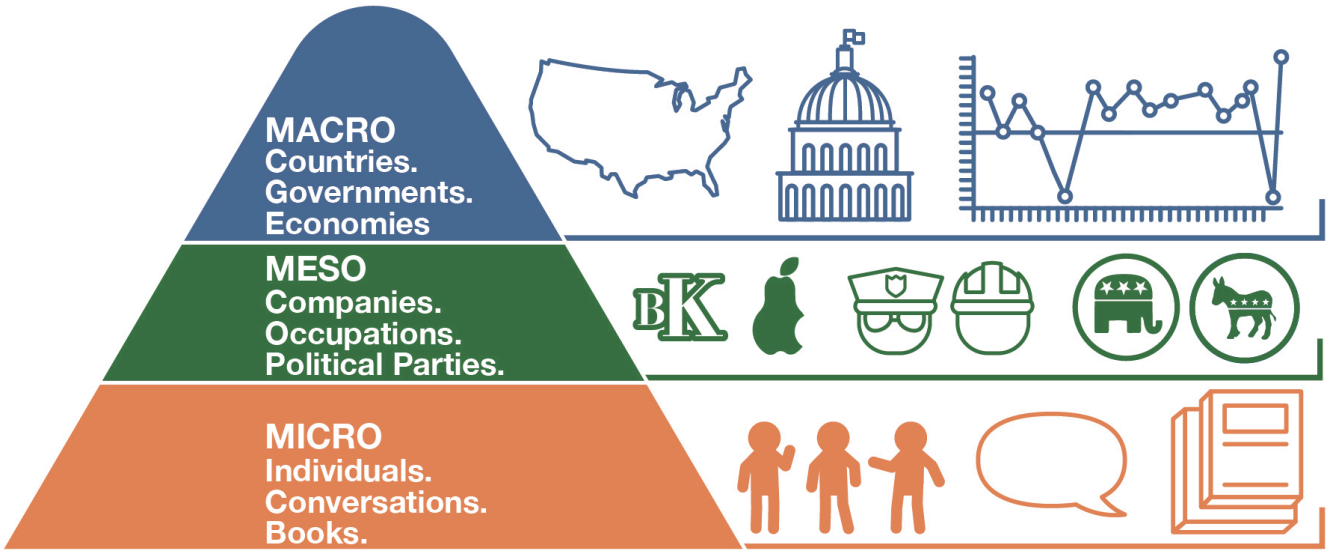


Figure 3.1. Macro, Meso, and Micro Levels of Analysis. Sociologists study phenomena at three levels of analysis: *macro*, *meso*, and *micro*. Typically, we're dealing with individuals (micro level), organizations (meso level), and societies (macro level), but as the examples in each section of the diagram show, there are many other possibilities.

Learning Objective

Define levels of analysis and how they influence our approach toward research.

In the hard sciences, researchers can study phenomena at very different levels—from stars and planets, to land masses and biomes, to microscopic cells and microbes, to the fundamental particles that constitute our material reality. In a similar way, sociologists study the social world using different **levels of analysis** (Figure 3.1). The three broad levels of analysis are *micro*, *meso*, and *macro*. We can think of this as a continuum from the lowest levels of analysis to the highest—from interactions between individuals, to comparisons of entire societies (Video 3.1).

At the **micro level**, sociologists examine the smallest levels of interaction—in some cases, just "the self" alone. Micro-level analyses might include one-on-one interactions between couples or friends. Or, perhaps a sociologist is interested in how a person's perception of self is influenced by their social context. In each of these cases, the level of inquiry is micro.

In addition to deciding on a level of analysis, sociologists also need to choose the units of analysis they wish to study (we'll discuss units of analysis in depth in Chapter 6: Sampling). A **unit of analysis** is what we are really studying—the class of phenomenon we ultimately wish to comprehend. Typical units of analysis include individuals,

groups, organizations, and countries. The unit of analysis at the micro level is often the individual, but that's not always the case. For example, when studying social interactions, we might focus not on individuals per se, but on **dyads**, pairs of people who are connected through that interaction (see the sidebar **Sociology across Boundaries** to learn about research on interracial couples). Or, we might focus on the interactions themselves—that is, making every conversation into a unit of analysis. This brings up another feature of units of analysis: they don't have to be people. We can analyze texts like books and magazines, visual content like paintings and TV shows, or events like protests and strikes (we'll talk more about some of these possibilities in **Chapter 15: Materials-Based Methods**). Our choice of a unit of analysis will depend on what we truly want to understand—individuals, or conversations, or books, or whatever.

When sociologists investigate groups, their inquiry is at the **meso level**. Sociologists who conduct meso-level research might study how norms of workplace behavior vary across professions (with specific occupations, like doctors or firefighters, being the unit of analysis), or how children's sporting clubs are organized (with a club being the unit of analysis).

At the **macro level**, sociologists examine social structures and institutions. Research at the macro level examines large-scale patterns. For instance, sociologists have long studied the process and impacts of *globalization*, the growing interconnectedness of the economies and cultures of the world. A research project that examined how nations are connected economically or culturally would be an example of a macro-level study. In this case, the unit of analysis would be a country, though sociologists can also design macro-level studies where the unit of analysis is larger—say, a region of the world—or smaller—say, cities or states in a particular country.



One or more interactive elements has been excluded from this version of the text. You can view them online here: https://viva.pressbooks.pub/sociology-research-methods/?p=101#oembed-1

Video 3.1. Macro and Micro Perspectives. This video illustrates the differences between micro- and macro-level approaches.

Note that the boundaries between micro, meso, and macro levels of analysis are relative and somewhat arbitrary. For example, some international nongovernmental organizations, like the International Monetary Fund or United Nations, represent many countries: when we study them, are we operating at a meso or macro level? Likewise, a quantitative analysis of poverty using state-level data might be considered a "macro" approach relative to ethnographic observations in poor neighborhoods. Using these labels helps us situate our research alongside other work, but we don't want to get hung up over the precise categories.

Let's take a closer look at some specific examples of sociological research to better understand each of the three levels of inquiry described previously. Some topics are best suited to be examined at one level, while other topics can be studied at each of the three different levels. The particular level of inquiry might shape a sociologist's questions about the topic. Sociologists might also view the topic from different angles depending on the level of analysis being employed.

Work by Stephen Marks offers an excellent example of research at the micro level. In one study, Marks and Shelley MacDermid (1996) draw from prior micro-level theories to empirically study how people balance their various life roles—for instance, their identities and activities as parents, partners, and workers. In this study, the researchers found that people who experience balance across their multiple roles report lower levels of depression and higher levels of self-esteem and well-being than their less "balanced" counterparts. In another study, Marks and his colleagues (2001) examined the conditions under which spouses feel the most balance across their roles. For women, having more paid

work hours and more couple time were crucial. For men, having leisure time with their families was important, but their "role balance" decreased as their work hours increased.

At the meso level, sociologists tend to study the experiences of groups or the interactions between groups. For example, Jason Spicer and Christa Lee-Chuvala (2021) studied a global network of banks dedicated to upholding social and environmental values. Tracking how membership in the network changed over time, they found that credit unions and other banks owned by their customers rather than investors were more likely to remain within the network and stay committed to their values. In a much different study of group-level data, Michael Messner (2009) conducted research on children's sports leagues. Messner studied interactions among parent volunteers, among youth participants, and between league organizers and parents. He found that the adults who run such leagues structure the children's gameplay in ways that reinforce gender boundaries and hierarchies.

Sociologists who conduct macro-level research study interactions at the broadest level, such as comparisons across cities or nations. For example, David Frank, Bayliss Camp, and Steven Boutcher (2010) examined changes in laws regulating sex across a number of countries between 1945 and 2005. They found that laws regulating rape, adultery, sodomy, and child sexual abuse shifted in focus from protecting families to protecting individuals. In another macro-level study, Leah Ruppanner (2010) studied how national levels of gender equality in 25 different countries affected how couples divvy up housework. Ruppanner found that as women's representation in a country's parliament increased, so, too, did how much time men in that country participated in housework. (Note that we don't always know what is driving the changes and associations we observe across these sorts of studies—in this case, how exactly greater women's representation in a country's government is linked to greater male participation in housework, or whether the first really causes the second. We'll have more to say about these issues later in this chapter.)



Sociologists can study a phenomenon like poverty from various levels of inquiry. From a bird's-eye view—the macro level—they might look at how poverty rates differ across states or even nations. Zooming in a bit closer, they could conduct a meso-level analysis of different neighborhoods within a city, examining their different approaches of supporting, tolerating, or repelling the poor. Equally valid, though, would be an in-depth study of poverty that looks at how individuals experience material deprivation in their day-to-day lives. For instance, in her classic micro-level study, Gwendolyn Dordick (1997) followed groups of homeless people in New York City, capturing in rich detail how they coped with conditions on the street, how they dealt with police and other authority figures, and how they fought and cared for one another. *Timur Weber, via Pexels*

Let's consider one important sociological topic—poverty—and then consider how researchers study it from each of the levels of inquiry just described. At the micro level, they might study what everyday life is like for people who are homeless by observing them where they panhandle and sleep. One example of a micro-level study of extreme poverty that uses this approach is Gwendolyn Dordick's book *Something Left to Lose* (1997), based on the author's intimate observations of the lives of homeless people in New York City. At the meso level, sociologists often study poverty by looking at issues that arise within poor neighborhoods—as Robert Sampson has done intensively in the city of Chicago, as described in his book *The Great American City* (2013). And at the macro level, they might compare poverty across communities within a nation—or even compare rates of poverty by country. David Brady and Amie Bostic (2015) used survey data from 37 countries to test earlier theories about the mix of social policies that were most effective in reducing poverty rates across those countries. They found that policies providing generous government assistance that was more universal—that is, not targeted just at the poor—were associated with the sharpest reductions in poverty.

While it is true that some topics lend themselves to a particular level of inquiry, many social phenomena can be

studied from any of the three levels. The choice depends on the specific interests of the researcher, the approach they would like to take, and the sorts of questions they want to be able to answer about the topic.

Sociology across Boundaries: A Q&A with Chinyere Osuji



Chinyere Osuji is a sociologist who studies how Africans and members of the African diaspora understand racial and ethnic "others." She examines the meanings they give to ethnic and racial differences as well as how they use other social categories to reinforce, blur, or erase those boundaries. Her first book, *Boundaries of Love* (2019), compared how Black-White couples in Brazil and the United States make sense of race. Currently, she is working on two book projects: one about African immigrants confronting race in the U.S. nursing profession, and another on anti-Blackness and White supremacy in South Korean media.

What drew you to study Brazil and to do your fieldwork in that country?

In graduate school, I found out that Brazil had the largest number of people of African descent in the world outside of Nigeria. I'd already had an interest in race in Latin America, and given the millions of people who identify as *negro*, Brazil became an obvious choice when trying to understand people of the Afro-Atlantic. Since I already spoke Spanish, I figured Portuguese would be easy to pick up.

In *Boundaries of Love*, you discuss the fluid and complex ways that people think about their racial and ethnic identities—not just contrasting the U.S. and Brazilian understandings of racial identity, but also talking about how multiple aspects of a person's identity and social context shape their personal experience of race. Can you talk about how your own understanding of race and ethnicity has evolved over the years, and how you came to the nuanced theoretical perspective you put forward in your book?

I grew up in Chicago in the 1980s. My parents didn't understand U.S. race relations. They'd come here in the late '70s from Nigeria with the idea that they were going to get a college education and then return home to the strong economy and newly minted democracy they'd left behind. But while they were living abroad, Western-backed coups and economic policies destabilized Nigeria and many other countries in the global south. Staying in the United States seemed like the best option.

Today, bilingualism is highly valued. However, when my parents arrived, the United States was informally an English-only society. My teachers encouraged my parents not to speak to me or my siblings in their native tongue, Igbo. They thought I would have trouble with English, even though my parents speak both languages fluently.

Our family lived on the north side of the city, where many recent immigrants from Africa tend to live. However, because we lived in public housing, we were surrounded by all sorts of low-income families, including Black Americans, who were the largest group in our building. There were also other families of African immigrants, Puerto Ricans, Mexicans, and even Eastern European refugees living there. Since Africans were their racial

status quo, my parents thought everyone with light skin—whether Puerto Rican, Middle Eastern, or Chinese—were *oyibo*: White people. Some Black American neighbors complained about our "stinky" African food. Meanwhile, African immigrants tsk-tsked the single-headed households of our Black and Puerto Rican low-income neighbors. One Afro-Cuban neighbor looked like us, but since she spoke only Spanish, I found her confusing—all the other foreign-born Blacks I had encountered spoke English.

Middle-class Black families lived in distant neighborhoods like Hyde Park or the south suburbs, so my understanding of Blackness from kindergarten to eighth grade was limited. It was only when I attended Whitney Young High School—also Michelle Obama's alma mater—that I saw Black Americans across class backgrounds. I learned that many behaviors that Americans—both Black and non-Black—had insisted to me were "Black" were really traits of the urban poor.

In college, I studied abroad in the southern Spanish city of Granada, where the old Moorish palace the Alhambra still stands. In my classes, I learned that Africans ruled over much of the Mediterranean long before the concept of race existed—sometimes in cooperation with Jewish and Christian brethren, sometimes not. Several people I met in Granada looked like what we might call "mixed" or "biracial" in the United States. There, I learned to speak Castilian Spanish, lisp and all. When native Spaniards tried to speak ill of me—thinking I was an African street vendor who could not understand them and deserved poor treatment—I replied in their own tongue. In Granada, I also witnessed discrimination against Spain's Roma population. Despite the fact that *gitanos* were seen as thieves and ne'er-do-wells, *gitano* culture was nationalized to the point that flamenco was seen as Spanish, not as a cultural product of their disenfranchised Roma people. Living in Spain—the birthplace of modern racism 550 years ago—made me interested in how people around the world understand ethnic and racial boundaries.

In your book, you push back against notions in popular culture that Brazil is a racially harmonious utopia. How did your work lead you to this view?

I came across several articles by Edward Telles, a future advisor and mentor of mine at UCLA. He had written a lot about racial equality in Brazil. According to his research, Afro-Brazilians earned less than their White peers even when they worked in the same occupations. They had lower levels of education even when they had parents of the same socioeconomic background. They also lived in neighborhoods that were racially segregated—even if less so than hyper-segregated U.S. cities. These racial inequities were not fully explained by class in the ways that many Brazilians had assumed.

When I first went to Brazil for my research, I lived in a wealthy White neighborhood in São Paulo. I soon realized that everywhere around me were White people who looked Italian or Portuguese. The only Black or brown-skinned people there worked as domestics, doormen, or day laborers. My landlady was strange, saying things questioning my intelligence or belittling me as a human. I found it odd how her olive-skinned daughter, who was close in age to me, avoided me. The only saving grace was my White roommate, Celda, who was really kind. She had grown up in one of Portugal's colonies in Africa, but African independence had forced her Portuguese colonizer parents to leave when she was young. Over drinks one night, Celda explained that our landlady had told her secretly that she did not like Black people and had advised her daughter not to spend time with me outside of the home. That was my cue to leave. At the end of the month, Celda helped me pack up all of my belongings. In the morning, I told our landlady that I did not like how she had been treating me and that I had decided to move out. She was completely gobsmacked as I left the keys on the table. This housing situation was when I first confronted the fraud of Brazilian racial harmony.

Your interviews with interracial couples about their past and present "romantic careers" delve into intimate territory: sexual preferences, racial and gender stereotypes, fraught family dynamics. How did you approach

your interviews on these topics, and how did you guard against the possibility of alienating or offending your interviewees?

There were a lot of uncomfortable moments as I was doing those interviews. I reminded myself that I was just listening to their stories. No matter how uncomfortable or unrelatable they could be, they were still their stories and deserved all the respect and care I could muster. Any analysis I did would not occur in the moment. I would take notes on my emotions on the ride back to my apartment in Rio de Janeiro or Los Angeles. If necessary, I would discuss interesting encounters with my roommate or my friends.

Through my interviews with couples, I learned that men often expect women to do the emotional labor of recounting details of their relationship. It didn't matter whether they were Black or White, or living in Los Angeles or Rio de Janeiro: wives crafted the narrative of their relationships, and husbands rubber-stamped them. Men's voices are important, too, and I wanted to give them space to share their own experiences of their interracial relationships outside of their wives' interpretations. For this reason, I conducted the rest of my interviews one-on-one—making sure that the spouse was not within hearing distance. This was more of a struggle in lower-income neighborhoods, where there was not a lot of space for interviewees to be isolated from other family members. I will never forget going to a shantytown to interview a man about his previous relationships. He quickly glanced from side to side and spoke in a hushed tone, looking very, very uncomfortable. The walls in his home were thin, and since we could hear his wife moving about in the next room, she could probably hear us as well.

What advice would you give an aspiring sociologist if they wanted to pursue the sort of cross-national interview-based research you do?

Do not do it until after you have finished your dissertation. My original goal was to study only Brazil, but I had a fellowship that required that I stay in Los Angeles for one more year. I thought, "Why not make this a comparative project?" I have no regrets, but I wouldn't recommend it to a graduate student. I'd recommend it to somebody who's already done work in one society or can collaborate with other researchers.

Key Takeaways

1. Sociological research can occur at micro, meso, or macro levels of analysis, ranging from the study of individuals and individual-level interactions, to the exploration of groups and organizations, to the analysis of entire countries and societies.
2. Some topics lend themselves to one particular analytical level, while others could be studied from any, or all, of the three levels of analysis.

Exercise

Look over the examples of research studies discussed in this section and see if you can figure out what the unit of analysis was. (You can go off the description in the next, or download the actual article if you have library access, as described in **Chapter 5: Research Design**.) For some of these studies, the unit of analysis will be obvious, but for others, you might make a case for one or the other—or even multiple units of analysis.

3.3. The Components of Theories

Learning Objective

Identify the different components of theories and what questions they answer.

Sociologists develop theories to explain social phenomena. At its most basic level, a theory is a proposed relationship between two or more concepts. Ultimately, a theory seeks to explain why a phenomenon occurs. A good theory must include four essential components (Whetten 1989):

- *What*? Which concepts should be considered as part of the explanation of the phenomena of interest?
- *How*? Now that we have identified the concepts, how are these related to each other?
- *Why*? What are the underlying dynamics that explain the proposed relationship between concepts?
- *Who*, *where*, *when*? What are the temporal and contextual boundaries of generalizability? In other words, what is the scope or range of the theory?

Let's discuss each of these components in turn.

What Are the Concepts We Are Studying?



In the social sciences, concepts like "masculinity" can have a wide range of meanings that depend on the particular culture or society you are studying. In American culture, for instance, rugged frontiersmen and lone cowboys would come to mind as traditional examples of what it means "to be a man." As with many social concepts, however, ideas of masculinity evolve, and today's crop of popular musicians, actors, and other public figures present younger generations with a very different set of masculine ideals. *Brett Sayles, via Pexels*

Concepts can be loosely defined as a mental image of a particular phenomenon that summarizes its key aspects. They are abstractions that allow us to organize and discuss our thoughts about reality. A concept captures what is distinctive or essential about that phenomenon—what applies universally, and not just to a single case.

For example, "masculinity" is a concept. What do you think of when you hear that word? Presumably, you imagine some set of behaviors, perhaps even a particular style of self-presentation. Through the upbringing we receive or the media we consume, we develop a set of shared understandings about our social world, including about what it means to be "masculine." Of course, we can't necessarily assume that everyone conjures up the same set of ideas or images when they hear this word. In fact, there are many possible ways to define the term. And while some definitions may be more common or have more support than others, there isn't one true, always-correct-in-all-settings definition. For instance, what counts as masculine may shift over time, from culture to culture, and even from individual to individual.

You might be asking yourself why researchers bother defining a term for which there is no single, correct definition. But this is true for any concept you might measure in a research study: there is never a single, always-correct definition. (This tends to be less of a problem in the natural sciences—remember when we bragged about how the social sciences were harder?) When we conduct empirical research, our terms mean only what we say they mean.

As researchers, we must be careful to explicitly and clearly define our underlying concepts. This stage of our thinking process is called conceptualization**.** When we read other people's work, we will want to pay close attention to how they conceptualized their work, too—which is just as important a part of their research as whether they competently conducted their interviews and analyzed their data. For instance, if we don't thoroughly understand how a researcher has defined their key concepts, we can't really understand the meaning of their research findings and conclusions.

Especially in quantitative research, conceptualization involves writing out clear, concise definitions for our key concepts. Think about what comes to mind when you read the term "poverty." How do you know poverty when you see it? Is it about how much money people make? What they can buy? Where they live? Whether they have enough to eat? How their standard of living compares to those of their peers? Perhaps we could define poverty as the state of being deprived of basic necessities of life. That seems like a reasonable place to start, and in the early stage of conceptualization, it's appropriate to brainstorm about the images and examples conjured up by concepts and play around with possible definitions.

This, however, is just the first step. It's also critical to consult previous work to understand how other scholars have already defined the concepts we're interested in. (We'll talk about finding and making sense of past studies in **Chapter 5: Research Design**.) This doesn't mean we must use existing definitions, but understanding how concepts have been defined in the past will give us an idea about how our conceptualizations compare with the predominant ones out there. It will also help us decide whether we should adopt, modify, or challenge those conceptualizations.

After we've identified a clear conceptualization that we're happy with, we should make sure that every term used in that definition will make sense to others. Are there any terms that need to be spelled out? If so, our conceptualization is not yet complete. Also, concepts possess different levels of abstraction. Some concepts, such as *weight* and *age*, are relatively precise and objective, while other concepts, such as *personality* and *prejudice*, may be more difficult to measure or visualize.

Sometimes, sociologists borrow concepts from other disciplines or even popular culture to explain a phenomenon of interest. For instance, *distance* is a basic concept of physics, but sociologists can apply this concept to understanding the degree of *social* separation between two individuals—whether they know the same people, or whether they feel affinity or hostility toward the other person's social background. As that example suggests, sociologists often use existing concepts that are familiar to many people as metaphors for social life. To take one example, Michèle Lamont (1994) described the ways that people within one group try to distinguish themselves from people outside that group as boundary work. In doing so, she took an everyday physical concept—boundaries—and used it to illustrate a social phenomenon in a more intuitive and accessible way.

Let's go back to the concept we started this chapter with, "masculinity." Suppose we decide to follow other researchers by defining "masculinity" as "the social roles, behaviors, and meanings prescribed for men in any given society at any one time" (Kimmel and Aronson 2004:503). So now we're done, right? Not so fast. If you've ever met more than one man in your life, you've probably noticed that they are not exactly the same, even if they live in the same society and at the same historical period. This could mean there are different **dimensions** of masculinity.

More complex concepts (here, the term *construct* is often used) encompass multiple ways of measuring or thinking about that particular concept. You might hear sociologists talk about **multidimensional** concepts**,** which are essentially overarching concepts formed from many underlying concepts. *Culture*, *creativity*, and *writing skill* are examples of these higher-order abstractions. For example, a person's *writing skill* could be measured by a person's *vocabulary*, *grammar*,

and *spelling*. Each of these dimensions are important when thinking of someone's writing skill, but each of them alone doesn't adequately capture the overall concept.

Concepts have multiple dimensions when we can think of ways that representatives of that concept—say, "masculine" individuals—differ in the ways they express or embody that same concept. For example, we might consider how masculinity varies by *region* (is masculinity defined differently in different regions of the same country?), *age* (is masculinity defined differently for people of different ages?), or *status* (is masculinity defined differently for people who are members of privileged groups?).

Even within one culture, masculinity can have a variety of manifestations or aspects. For example, the Bem Sex Role Inventory (Bem 1974) includes multiple dimensions to describe traditional American masculinity, such as individualism, athleticism, dominance, aggressiveness, self-reliance, willingness to take risks, and leadership ability. Depending on your research question, each of these dimensions can be measured separately, or they can be combined into a single measure assessing overall masculinity.

Defining our concepts precisely and capturing all their relevant dimensions becomes even more important once we start measuring them in the real world. For instance, a seemingly simple concept such as *income* can be measured in many different ways. Are we talking about monthly income, or annual income? Their income before taxes, or after taxes? Their personal income, or their family income? Scientific research requires **operational definitions** that define concepts in terms of how we will actually measure them when collecting our data.

Stating how a concept will be measured—that is, defining it operationally—is called **operationalization**. We use a particular term to describe the operationalized form of a concept: **variable**. At its most basic level, a variable is a quantity or characteristic that can vary, and scientists often use this term interchangeably with "concept." Formally speaking, however, a variable refers to a measurable representation of a concept—in other words, one particular way (out of many possible ways) that we have decided to measure this intangible idea. As abstract entities, concepts are not directly measurable. Variables are the ways that we *choose* (remember the idea of choice here) to measure them according to some operational definition. We'll describe in detail the process of developing conceptual and operational definitions for our concepts in **Chapter 7: Measuring the Social World**.

If we are collecting our data for a quantitative research project, we generally need to do a lot of the work of conceptualization and operationalization upfront. Among other things, we should be certain about how we've defined our concepts before we start fielding the **research instrument**—that is, the survey questionnaire or interview guide we're using to measure those concepts. The content of the instrument must reflect the conceptualization process. For example, if we are surveying people about what policies they think should be enacted to reduce "poverty," we'd need a clear definition of the term to guide us as we write questions for our questionnaire.

Conceptualization in qualitative research proceeds a bit differently than in quantitative research. Because qualitative researchers are usually interested in the understandings and experiences of their **research participants**—the people or communities they study—it is less urgent for them to settle on fixed conceptual definitions before starting to interview or interact with participants. For instance, if you were conducting qualitative research on the meaning of poverty, you would likely start by consulting previous literature and coming up with a **working definition** for poverty. This working definition would merely be a place to start, and you would not think of it as the only or best definition out there. (Working definitions often change over the course of your investigation, particularly as you get to know your field site or the group of respondents you're interviewing.) You would bring this working definition of poverty into the field with you, but the definition that would really matter would likely be the one that your participants shared through their comments and behaviors during your data collection.

Note that your research participants' definitions of a concept might not fit well with how scholars understand it. Consider, for instance, how differently ordinary people think about racial categories than researchers who study race

do. These popular, or folk, understandings of concepts may or may not be "right," but for the qualitative researcher, what's often more important is how a person's definition of a term may influence the ways they act. In this regard, your participants serve as indispensable experts that guide your research, telling you how people from their group or background understand any concepts you uncover. Typically, your job here as the researcher is to accurately and reliably record and interpret your participants' understanding of the concepts they describe—not to impose your own understanding on them.

Because qualitative concepts change as researchers learn more information from their participants, getting them to comment on, extend, or challenge the definitions and understandings of other participants is a hallmark of qualitative research. This is typically the opposite of the approach taken by quantitative researchers. For quantitative studies, definitions must be set in stone before the data collection can begin. That's because the research instruments—the survey questionnaires being used, for instance—can't usually be revised after you start using them. Otherwise, the data you collect before and after those revisions can't be compared on an apples-to-apples basis.

That said, regardless of whether you have chosen qualitative or quantitative methods, you should have clear definitions for your chosen concepts, and you should make sure that all the terms you use in those definitions are equally clear. That way, other scientists will know exactly what your terms do and do not mean—and, as a result, can evaluate your work fairly and accurately.

Deeper Dive: Revising Our Social Realities



Our definitions of everyday concepts change with each new generation. In earlier times, popular understandings of what a "family" was would have been comparatively narrow—say, a married man and woman with children. Today, other types of households are becoming more common, and the "typical" family is much harder to pin down. When a society's perception of a concept evolves, sociologists may also need to define and operationalize it in new ways. *Kampus Production, via Pexels*

Conceptualization must be open to revisions—even radical revisions—as scientific knowledge progresses and cultures change. We should remember that concepts are abstractions from reality, but they are not themselves reality. Forgetting this fact sometimes causes problems for researchers, and that's why we have a special name for this error in logic: reification. **Reification** is the assumption that an abstract concept exists in some concrete, tangible way. If we wanted to be a true skeptic, we might see elements of reification in all aspects of our social lives. For instance, what is a "nation" like the United States, China, or India? Can we say that a "nation" has certain goals or preferences? If you peel away at this problem, you'll find that at the end of the day the concept breaks down, and what we're really talking about is specific government officials, or specific segments of the country's population, who feel and act a certain way.

As social scientists, we can go down a rabbit hole of questioning everything and anything, and as we'll describe

later in this chapter, some sociologists adopt this critical perspective. But even if you acknowledge how simplified our social reality can be, you still want to be wary of how reification can muddy your thinking as a researcher. Think about the term "family." If you were studying this concept, it would be important for you to consider the many differing ways that scholars—and the broader public—have conceptualized "family" over the years. Fifty years ago, researchers would have understood families in narrower ways than is common today. For example, they might have defined a "family" as "people related to one another by blood, marriage, or adoption." Their research on families would therefore have considered only those social groups fitting that definition. But if we adopted such a definition today, it would leave out, among other things, the growing numbers of people who choose to live together without marrying. In this case, we would want to avoid reifying the concept of family by taking it to eternally mean one sort of household arrangement.

It's not that researchers in the past were wholly wrong to adopt definitions of concepts like "family" that researchers today have largely abandoned. They were basing their definitions on what people knew at the time, and what practices were common. Their definitions of family were as real to them as our definitions are to us today. Furthermore, scholars typically want to reach some sort of lasting agreement about what various concepts mean, so that they can compare their results across studies and time periods. That said, if each new generation of researchers never challenged past conceptualizations, our scientific knowledge would be filled with blind spots, anachronisms, and prejudices from years ago. As important as scientific consensus is in building a robust body of knowledge, sociologists should never forget that particular people—from particular cultures in particular historical moments—have assigned the definitions widely used to describe the social world. These definitions are imperfect and subject to change.

How Are Our Concepts Related to Each Other?



Sociologists often want to identify patterns in the relationship between key concepts. For instance, a long line of sociological research examines how people's social class (concept A) affects their approach to parenting (concept B). As we describe next, these potential associations (say, between concept A and concept B) should be mapped out to guide our approach to studying them. An enterprising sociologist will find interesting combinations of concepts to study, measure how they relate to each other, and describe the specific processes that bring about noteworthy outcomes. *Ketut Subiyanto, via Pexels*

To explain social life, sociologists don't just identify key concepts relevant to a given social phenomenon. They also show how those concepts are related to each other. This can be as simple as saying that one type of group is associated with one type of behavior. For example, for her classic ethnography *Unequal Childhoods* (2011), Annette Lareau and her research team observed how parents raised their children. She found that middle-class parents took on a different parenting strategy than working-class or poor parents. Middle-class parents were more likely to act in ways that we associate today with "helicopter parenting"—filling their children's schedules with enrichment activities, taking an active role in shaping academic and extracurricular interests, frequently engaging with their kids to talk through intellectual topics. Less affluent parents did not love or care about their children any less, Lareau argued, but they left their kids to their own devices, reasoning that children should be allowed to be children, with freedom and unstructured time to spend as desired. In Lareau's theory of parenting, being a middle-class parent was associated with more engaged parenting—what she called *concerted cultivation*—and being a working-class or poor parent was associated with more hands-off parenting—what she called the *achievement of natural growth.* (Like Michèle Lamont did with her theory of boundary work, Lareau uses familiar imagery to make her theory more accessible—here, invoking a gardening metaphor of tending to plants.)

Depending on their intended use, variables may be classified as independent or dependent variables. An **independent variable** is a variable that we believe *explains* another variable. Specifically, changes in the independent variable are thought to cause changes in the other variable, which we call a **dependent variable** (because it "depends on"—*is explained by*—the independent variable). In the example study we just described, *social class* was Lareau's independent variable, and *parenting strategy* was her dependent variable: being from a particular social class meant that parents raised their children in a particular way, Lareau argued.

Let's break down how independent and dependent variables relate to one another. First, it's worth mentioning at the outset that the letters "X" and "Y"—either in their uppercase or lowercase forms—are traditionally used to designate independent and dependent variables, respectively. (When scientists create graphs to show the relationships between two variables, they typically represent the independent variable with the horizontal X-axis and the dependent variable with the vertical Y-axis.) Second, it sometimes helps to use analogies to think about the relationship between your independent and dependent variables. For instance, think of a speaker whose volume you control with a knob. The dependent variable in this (contrived) case would be the volume. Here, we are measuring (in our lingo, operationalizing) volume in decibels, and our specific measurement, or **value**, of the variable can range from zero (no sound) to however loud the speaker can go. The setting of the knob is our independent variable: if the knob went from 1 to 11, the value of our independent variable would range from 1 to 11. We can manipulate the setting for our independent variable (that's why it's "independent") and then the volume of the speaker will "depend" on the knob's setting. If the values of the dependent variable change as we shift between values of the independent variable, we say these two variables

are *correlated*, or associated; if the values of the dependent variable are unaffected, we say the two variables are *uncorrelated*, or unassociated. **Correlation** means that the two variables are related to some extent—an extent that we can actually measure with quantitative analysis, as we will see.

The example we just gave used two variables that were measured numerically—volume setting and decibels. But variables don't have to be quantitative. As we will describe in **Chapter 7: Measuring the Social World**, concepts like gender can be categorized qualitatively—varying across categories like male, female, nonbinary, and so on (these different categories are called **attributes**). A qualitative variable can serve as an independent variable, with shifts across its categories possibly explaining another variable. It can also be a dependent variable, with changes in another variable helping determine which of its values or attributes we see.

When we are analyzing concepts that can be quantified—whose measured values (i.e., variables) can increase or decrease—we can go further in describing the relationship between them. Specifically, we can describe any correlation that exists as positive or negative—what social scientists call the **direction of the relationship**. A **positive relationship** between an independent variable and dependent variable means that as the value of one variable increases, the value of the other variable also increases. A **negative relationship** (or *inverse relationship*) between two variables means that as the value of one increases, the value of the other decreases, and vice versa. Remember that this idea of "directionality" applies only to numerical variables that can actually increase or decrease.

Let's use the work of political scientist Robert Putnam to illustrate these two relationships. Putnam's seminal book *Bowling Alone* (2001) argued that civic engagement was on the decline. His analysis found that Americans' involvement in civic life—how often they volunteered for community organizations or joined clubs, whether they voted, how active they were in their religious communities, even how often they bowled in bowling leagues—had fallen dramatically over the previous decades. Although a number of factors contributed to this drop in engagement, Putnam highlighted the role of television. To put it in simple terms, his theory proposed that *the more TV people watched, the lower their involvement in civic life.* And his data appeared to back up that claim, with measures of civic engagement falling as television watching increased in the late twentieth century.



Drawing from survey data, the political scientist Robert D. Putnam put forward a frightening theory—challenged by some sociologists and other scholars—that television had led to a dramatic decline in civic engagement in the late twentieth century. *cottonbro studio, via Pexels*

Here, we can see how Putnam's theory contains the *what* and *how* elements of good theories. It proposes a relationship between two or more concepts—in this case, civic engagement and television watching. It describes the direction of the relationship, which happens to be negative: as civic engagement goes up, television watching goes down, and vice versa. Putnam's theory also put forward specific reasons that television watching might be causing lower civic engagement—the *why* element of theory. In his view, people were using up more of their limited free time to watch television, thanks to how enticing and pervasive this content had become. As a result, they didn't have the free time to get involved in their communities. Note that Putnam didn't have as much evidence to back up this *why* part of his theory—that television watching *caused* people to become less engaged citizens. He saw a clear correlation between these two variables, with their measured values moving in step with each other over the years, but it was harder for him to make a definitive case that television watching actually *caused* lower involvement. Specifically, Putnam couldn't

rule out that people were just less interested in volunteering for reasons that had nothing to do with the appeal of television. (We'll talk more about this issue of causality later in the chapter.)

When describing the relationships between their concepts of interests, qualitative researchers tend to shy away from using terms like "positive correlation" or "negative correlation," which fit more with the formal language of quantitative research. That said, we can still talk about the direction of many relationships in qualitative studies. Take Lareau's study of parenting. We could quantify her two concepts—saying that middle-class households are higher up on a scale of socioeconomic status, and saying that concerted cultivation is the state of being more engaged as a parent. Then we could say that a positive correlation exists between socioeconomic status and parental engagement: as socioeconomic status rises, parental engagement increases. Note that qualitative scholars may protest when other researchers reduce the complex relationships they describe into simple quantitative associations. We'll talk more about these issues in **Chapter 11: Qualitative Data Analysis**.

Especially in quantitative studies, researchers will explicitly state the relationship that they believe exists between two or more concepts in a **hypothesis**.[1] The hypothesis is a scientific conjecture—an educated guess—about how these concepts are related. Here's an example of a hypothesis: "An increase in a child's family income causes an increase in the amount of education they later attain." This statement does not have to be true, but it must be empirically testable—that is, the researcher must be able to use data to judge whether it is true or false.

A good hypothesis also specifies how the underlying concepts will be measured. We can make our example hypothesis more concrete by using specific variables—specific operational definitions—in place of its concepts. For example, since annual household income and years of education are ways of measuring income and educational attainment, respectively, we can substitute them in our hypothesis, like so: "An increase in the annual household income of a child's family causes an increase in the number of years of education they later attain."

How do we develop hypotheses? We can generate them using logic. Given what we know about existing theories, what sorts of reasonable and concrete hypotheses can we think up, which we can then go out and test? Our empirical observations (or those of past studies) may also suggest certain patterns—certain relationships between concepts—and our research can see if those patterns hold up more widely. (As we will describe in the next chapter, these two strategies—one starting with theory and one starting with observations—can be described as a **deductive approach** and an **inductive approach**, respectively.)

Let's connect what we've learned so far regarding concepts, variables, and hypotheses. **Figure 3.2** illustrates how social scientists operate on the two levels we described at the beginning of this chapter: a theoretical level and an empirical level. Through theory, they connect abstract concepts that describe our social world, and through empirical observation, they measure those concepts using concrete variables. They generate hypotheses that link multiple concepts—and link the variables corresponding to those concepts—and put forward their best guesses about the nature of those relationships. Then they test those hypotheses using the observations they have systematically collected.

Thinking like a researcher means becoming skilled at moving deftly back and forth between these theoretical and empirical levels. As social scientists, we ultimately are interested in theoretical relationships—the ways that abstract concepts like income and education relate to one another and thereby influence lives. But we conduct research in the real world, which means we need to measure these concepts in concrete terms—by dollars of annual household income, for example, and years of acquired education.

1. Sometimes, scientists distinguish between *propositions*—conjectures about the relationships between concepts—and *hypotheses*—conjectures about the relationships between variables (Bhattacherjee 2012). (Remember that *variables* are technically the measurable representations of abstract *concepts*.)

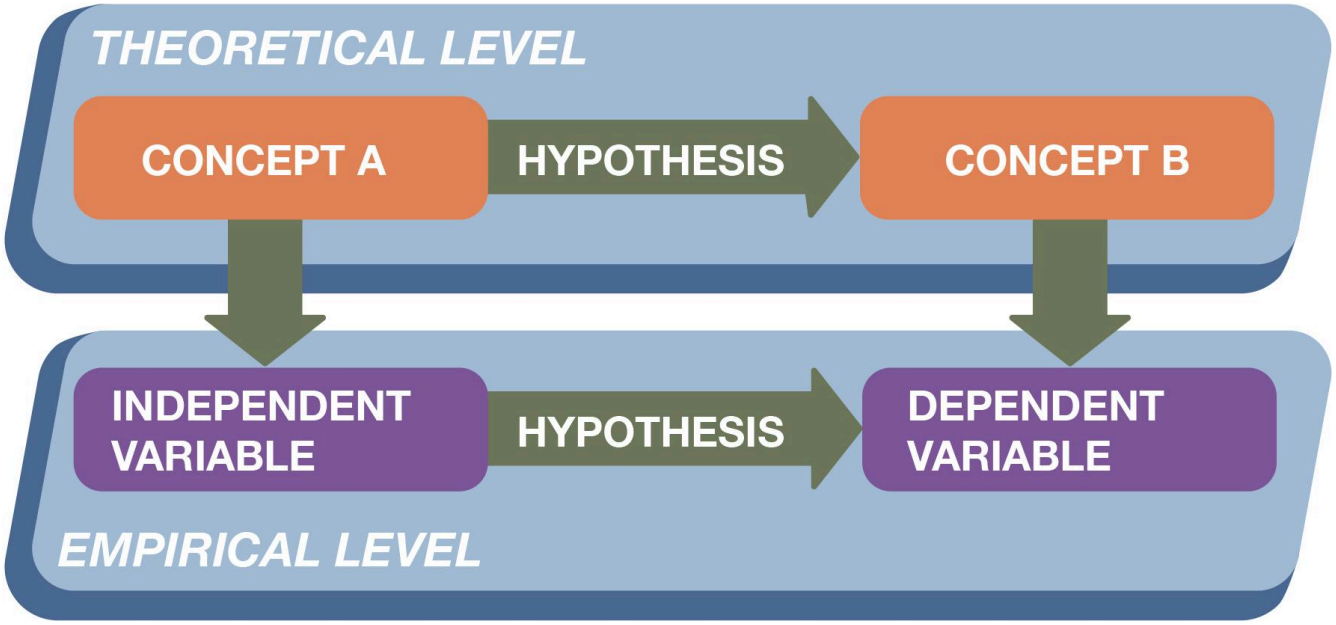


Figure 3.2. The Theoretical and Empirical Levels of Research. This diagram shows the distinction between concepts (conceptualized at the theoretical level) and variables (measured at the empirical level), and how we propose the existence of relationships between them by generating hypotheses.

As you consider how concepts relate to one another, it's useful to visualize those relationships with boxes and arrows, as shown in **Figure 3.3**. This is called a **concept map**. Each box refers to a specific concept (or its measurable representation, a variable). The plus or minus sign above the arrow in each diagram indicates the direction of the relationship: positive or negative. (Note that the relationship between qualitative concepts won't have directionality, as they can't be said to "increase" or "decrease." For example, more men than women may say blue is their favorite color, but we can't say this relationship is positive or negative, because neither concept—gender or color—is being quantified.)

The arrows between the boxes show the **direction of causality**. "Direction" in this sense is different from the "direction" of a relationship. In the first two diagrams, the arrow goes from concept A to concept B, and so we say that a change in concept A *causes* a change (either positive or negative) in concept B. If the arrow were reversed, we would say a change in concept B causes a change in concept A. Sometimes the direction of causality is obvious. Hot summer days cause more ice cream consumption, as any ice cream van driver knows well, but your decision to eat gallons and gallons of rocky road ice cream will have no effect on local temperatures (unless you eat so much that global ice-cream production must ramp up to satisfy your insatiable appetite, thereby generating more carbon emissions and worsening climate change—not really the likeliest scenario, but we won't judge). In many cases, however, the causal arrow could go in either direction, as we will discuss in more detail.

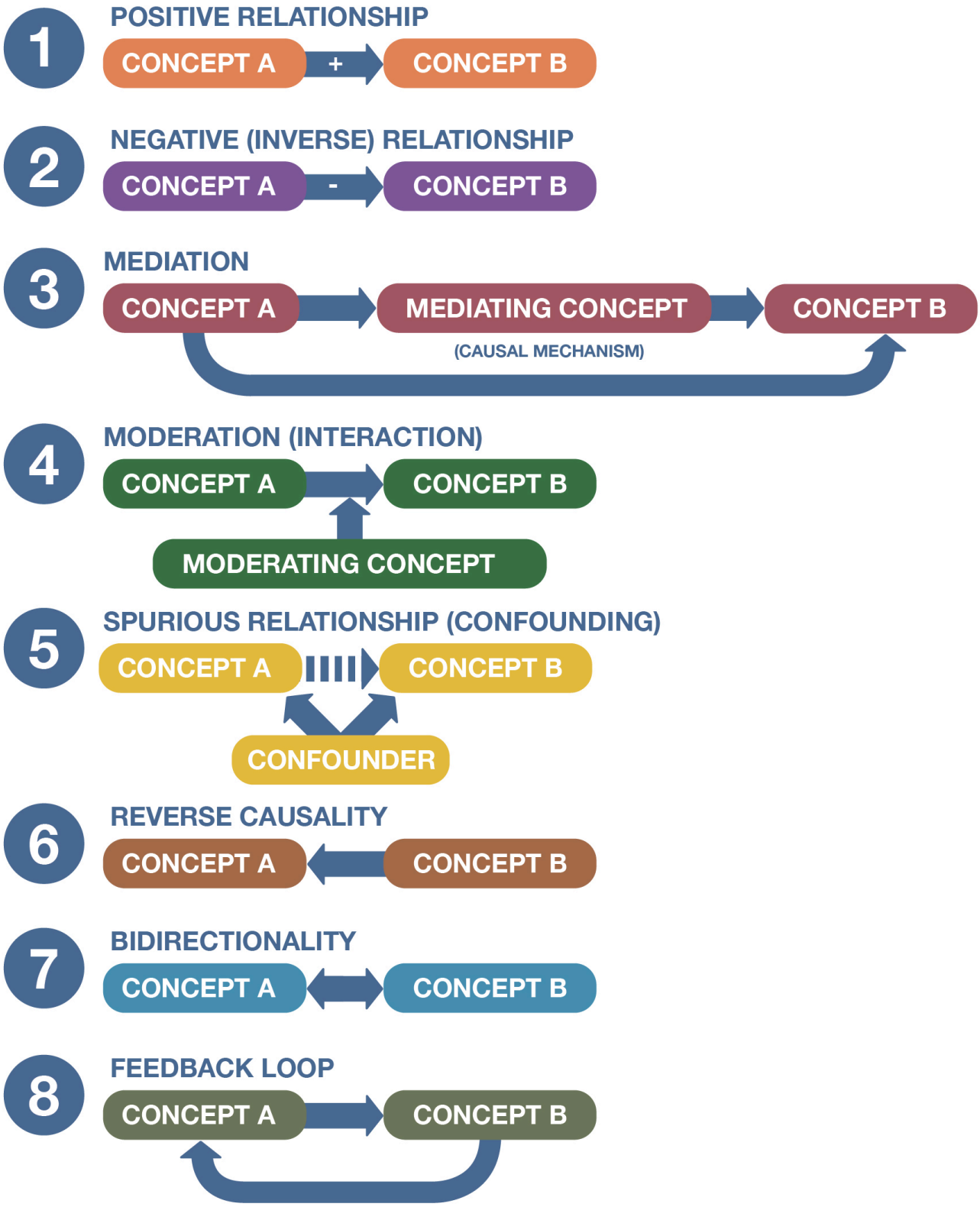


Figure 3.3. The Relationships between Concepts. These diagrams show different types of relationships that can exist between two or more concepts.

Why Are Our Concepts Related to Each Other?

As we have described, theory is the tissue that connects raw data and critical thought. To put this another way, the *what* (concepts) and *how* (relations between concepts) constitute the subject of the theory—what we are studying—but they are merely descriptive. The *why* provides the explanation. Data (be it quantitative or qualitative) describes or characterizes; theory explains.

In *Bowling Alone*, Putnam analyzed data over decades showing that civic engagement had declined and television watching had increased. But data alone is not particularly informative. If Putnam had not proposed a relationship between the two elements of social life, we may not have realized that television viewing does, in fact, reduce people's ability and desire to participate in civic life. To understand the social world around us, it is necessary to develop theory to draw the connections between seemingly disparate concepts.

Another example of sociological theorizing illustrates this point. In his classic work *Suicide* (1897), Émile Durkheim was interested in explaining a social phenomenon, suicide, and employed both data and theory to offer an explanation. By aggregating data for large groups of people in Europe, Durkheim was able to discern patterns in suicide rates and connect those patterns with another concept, religious affiliation. Durkheim found that Protestants were more likely than Catholics to die by suicide. But why exactly was this the case? To explain this relationship, Durkheim applied his theory of social solidarity, which claimed that not being closely connected to other people—which he argued characterized life in many industrialized societies—would make individuals feel a lack of direction and stability, putting them at greater risk of suicidal thinking. Applying his broader theory to the data he was analyzing, Durkheim argued that the looser social ties found in Protestant religions led to weaker social cohesion, which in turn made Protestants more likely to die by suicide.

Without good explanations for why different concepts are related, we cannot understand what factors cause what outcomes in social life. Analyzing these cause-and-effect relationships is a key way that the social sciences help us to better comprehend our reality. While Durkheim's analysis was flawed (a topic we'll explore further in **Chapter 7: Measuring the Social World**), his study was a novel and pioneering use of theory to explain the relationship between two concepts that we might normally think have little to do with one another—suicide and religious affiliation.



You can think of the independent variable (or more broadly, the concept it represents) as a button or crank on a machine. As the value of that variable changes, the output of the machine—the dependent variable—changes as well. But what is happening within the machine? We can think of the processes happening "inside" as *causal mechanisms*, which connect the independent and dependent variables in a similar way to how a machine's gears, pulleys, switches, and other parts convert the operator's input into the machine's output. Qualitative work in particular is good at identifying and understanding those mechanisms, whereas quantitative work tends to specialize in measuring the impact of particular variables on that output—including the impact of *mediators* (also called *mediating concepts* or *mediating variables*), which amount to another way of understanding causal mechanisms. *Tima Miroshnichenko, via Pexels*

Let's return to **Figure 3.3**. Diagrams 3 through 5 can help us flesh out our thinking about how and why concepts might be related to one another. So far we've been talking about relationships between two concepts, but as these diagrams illustrate, theories can link three or more concepts. Here you can see representations of "mediated," "moderated," and "spurious" relationships. You should know that these terms, like the terms "positive" and "negative" correlations, are more frequently associated with quantitative methods—even though qualitative researchers must also think carefully through the complex ways that their concepts might be related to another. It's worth reiterating that many social scientists will use the term "variable" instead of "concept" when discussing these relationship types—they'll talk about "mediating variables," "moderating variables," and "confounding variables," for instance. We're going to use the term "concept" to be precise, because variables are really the operationalized form of the underlying concepts, but many studies you read will do otherwise.

In **mediation**, a concept essentially stands between and links two other concepts. You can think of a *mediating concept* (also known as a **mediator** or **linking concept**) as the pathway by which one concept affects another. As you can see in **Figure 3.3, Diagram 3**, the mediating concept is *explained by* concept A—that is, the measured value of the mediating concept changes as the value of concept A changes. In turn, concept B is explained by the mediating concept—its values change as the values of the mediating concept changes. If the degree of mediation is particularly strong, the linking concept can largely account for any influence that concept A has on concept B—it's the **causal mechanism** here, the specific way that concept A affects concept B.

Let's consider an example. Involuntarily losing their job often makes people feel depressed or anxious. It does so through specific mechanisms—for instance, by making people scared about whether they can pay their bills and otherwise make ends meet (what's called *financial strain*). We could therefore say financial strain is an important mediator between unemployment and psychological distress. One way that quantitative researchers identify linking concepts like financial strain is by seeing if the correlation they observe between two concepts disappears when a third concept that logically could be driving that relationship is accounted for in their analysis—in that case, the third variable may play a mediating role.

When we discuss qualitative research in later chapters, we'll have a lot to say about mediating concepts, which qualitative researchers are particularly good at identifying. Qualitative researchers, however, tend not to use the

term "mediation," but instead often talk about causal mechanisms (also called *social mechanisms*)—again, the specific ways that one concept affects another. You can think of social mechanisms as the literal "mechanisms"—the gears and belts—within a machine. With a physical machine, you put something into the box—say, harvested cotton in a cotton gin—and out comes something else—the cotton fibers stripped of their seeds. But what is happening within the machine? What mechanisms are operating that turn the input into the output? These are the causal mechanisms we wish to identify within the intangible "black box" of social life. They get at the essential *why* questions we seek to answer with our theories.

Indeed, qualitative studies frequently focus on these causal mechanisms, trying to flesh out a **causal story** (or *explanatory story*) of how exactly the input leads to the output. For example, countless studies have investigated how coming from a wealthy family makes it easier to become successful. But how exactly does that occur? Is it based largely on how much money one's parents can pay for important investments like college tuition, or is it based on other resources that the family can bring to bear, like job contacts they have access to, or the knowledge and habits they instill regarding how to present oneself as a professional? Here, the ways that qualitative methods generate rich detail and illustrate complex processes can give us a particularly insightful understanding of the causal mechanisms linking concepts. Take Shamus Rahman Khan's (2011) influential study of an elite private school, which showed how children of wealthy families benefited not just from access to costly academic opportunities, but also from less tangible advantages—such as the outward sophistication and "ease" they had acquired from their upbringing, which helped them win over college admissions committees and other important gatekeepers. Through studies like Khan's, we can see the many concrete ways that the input of family wealth brings about an output of later success. Qualitative researchers focus so much on causal mechanisms in part because of their inclination to explain processes in such detail—in a sense, deciphering the machinery of the social mechanisms at work.

In **moderation** (also known as *interaction*)**,** a concept influences or shapes the relationship between two concepts. Specifically, the presence of this moderating, or *conditioning*, concept weakens or strengthens (or otherwise affects) the relationship between concepts A and B. As you might guess, it's very easy to confuse mediation and moderation, given the similarities in what they do. It may be helpful to think of mediating concepts as *standing between* two concepts (**Figure 3.3, Diagram 3**)—that is, being the pathway by which one concept affects the other—and moderating concepts as *changing* that relationship (**Figure 3.3, Diagram 4**).

Let's consider an example of a moderating concept. Katharine Donato and Gabriela León-Pérez (2017) analyzed data collected in Colombia to investigate if individuals with more education were more likely to emigrate to the United States. Indeed, the data showed that higher educational attainment was associated with a higher likelihood of migration. However, this relationship operated differently for men and women: as education increased, the corresponding boost in the likelihood of migration was greater for women than for men. We might say that gender is a moderating concept that changes the relationship between education (concept A) and migration to the United States (concept B). Put in a different way, a *moderation effect* (or *interaction effect*) is at work here: the association between an individual's education level and their migration decisions is different for men and women, given how gender moderates that relationship, or how gender and education interact in predicting migration.[2]

2. Technically, we would use the term "moderation" when we are clearly interested in how two concepts relate, and we are interested in a third concept only insofar as it moderates, or conditions, the relationship between those first two concepts. If we are not so focused on the relationship between two particular concepts, then we wouldn't be so specific in our language. We'd say an interaction existed between two concepts in predicting the values of a third concept, but we wouldn't single out one of those concepts as a moderator. The distinction here is similar to the one that we make between correlation analysis and regression analysis, which we'll cover in **Chapter 14: Quantitative Data Analysis**; for

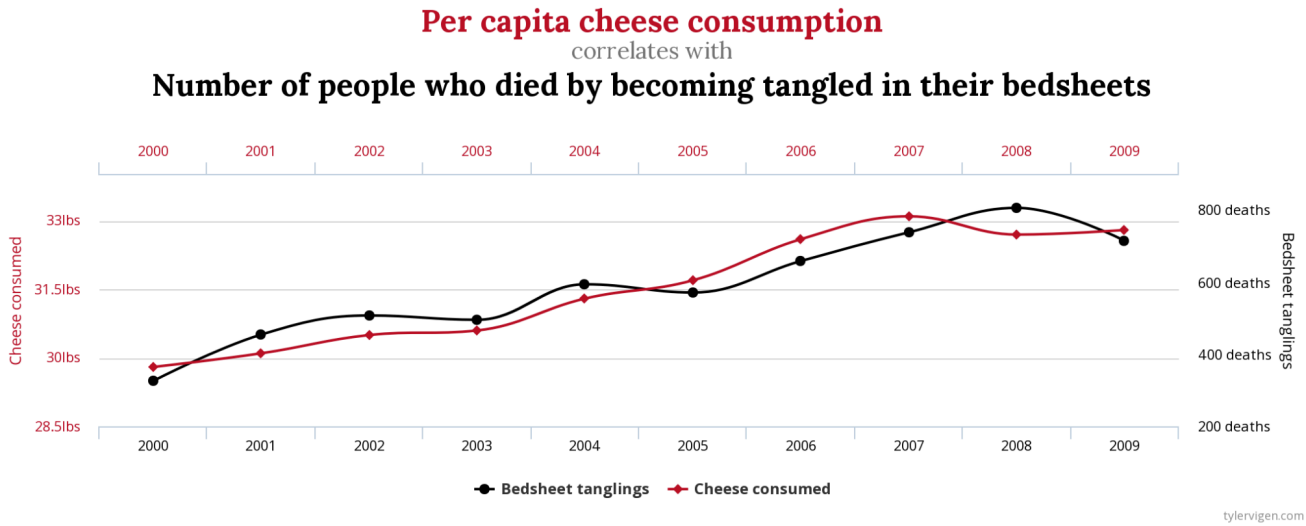
For quantitative variables, you might think of a moderating variable as either (1) steepening or (2) flattening the slope of the line that describes the relationship between two variables. In the first case, it makes the presumed effect of the first variable on the second stronger. In the second case, it makes the presumed effect weaker. Sometimes, in fact, moderation can turn a positive relationship into a negative one, and vice versa. (We'll talk more about how relationships between variables can be graphed in **Chapter 14: Quantitative Data Analysis**.)

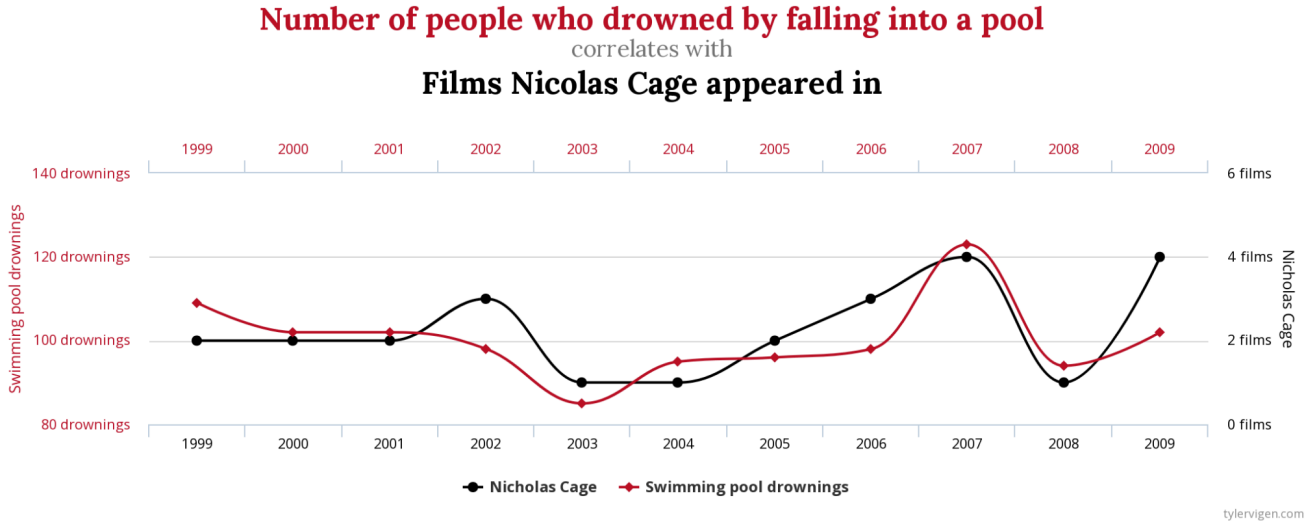
Finally, **spuriousness** refers to situations in which the relationship between two concepts seems to exist but, in reality, they are linked by a third concept, what is called a **confounder** (or *confounding variable*). **Figure 3.3, Diagram 5**, depicts a spurious relationship. In this scenario, concepts A and B are correlated, but no actual causal relationship exists between them (hence the red X). Instead, the confounder is related to both concepts; it "confounds" our understanding of the actual relationship between them. One well-known example of a spurious relationship is the link between coffee consumption and mortality: people who drink a lot of coffee have, on average, a higher risk of death. But a confounding variable—smoking—explains this relationship. People who smoke tend to drink coffee, and therefore the higher mortality rate we observe for coffee drinkers is due to the fact that they disproportionately smoke and are vulnerable to the many illnesses associated with smoking. In fact, when researchers statistically account for tobacco-smoking and other confounders, they find an *inverse* relationship between coffee consumption and mortality—that is, more coffee drinking is associated with a lower, not greater, risk of death (Freedman et al. 2012). (We call a confounder that actually reverses the association between the independent and dependent variable a **suppressor variable**.)

Let's consider another example of potential confounding, which we'll return to throughout this section. We may see a correlation between the amount of time children play violent video games and the aggressive behaviors they display, but that doesn't mean the relationship is causal (this is a contentious issue in the research literature on this topic, in fact). Instead a third variable—say, a lack of parental attention—may cause both the playing of violent video games (perhaps because children are left unsupervised) and the aggression seen in children (perhaps because the children feel unloved).

You might have heard the saying, "Correlation is not causation," and it is advice to take to heart: just because we see a correlation between two concepts doesn't mean the relationship is causal. In the policy world, it is especially vital to identify spurious relationships because we can draw erroneous conclusions based on them. People feel strongly about banning or regulating certain products—from violent video games, to pornography, to guns—based on the correlations observed between their use and various negative social outcomes. However, it is not always clear that the relationships are causal—we just mentioned one possible confounder when examining the relationship between violent video games and aggression. If the underlying relationship is not causal, then a proposed ban or regulation won't bring about the positive benefits we hoped to see. Instead, we will need to identify and address the *real* cause of the problem—which may happen to be a confounder we overlooked or couldn't easily measure. Clearly, we don't want to be making policies based on spurious correlations. We may spend lots of time, money, and other resources trying to influence one factor when another is really driving the phenomenon we're interested in.

regression, we specify an independent variable and dependent variable, but for correlation, we are simply observing associations between variables, with no cause-effect relationship between those variables in mind.





In his book *Spurious Correlations* (2015), Tyler Vigen unearths startling evidence of the empirical relationships between horrific outcomes and some seemingly ordinary variables. For instance, did you know that how much cheese the average person consumes is highly correlated with how many people die every year from getting tangled in their bedsheets? Or that the number of people who die by falling into a pool tracks closely with the number of movies that actor Nicholas Cage appears in over the course of a year? In fact, if we measure these two relationships using statistical methods (discussed further in **Chapter 14: Quantitative Data Analysis**), we obtain correlation coefficients of 0.947 and 0.666, respectively (where 0 is no correlation and 1 is perfect correlation). Read below for more details about these deeply troubling findings. *Tyler Vigen, via Spurious Correlations*

We can point to two possible reasons for spurious relationships. First, the relationship we observed may simply be due to chance. A website (and now book) called "Spurious Correlations" describes a number of fortuitous relationships between bizarrely matched variables. For example, it turns out the number of films the actor Nicholas Cage appears in every year is highly correlated with the annual death toll from people drowning by falling into pools. The correlation is even stronger between annual per capita cheese consumption and the number of people who die each year by becoming tangled in their bedsheets. Of course, none of these variables are related to the others. But sometimes social scientists can be fooled into thinking a causal relationship exists between two variables because changes in one appear to follow changes in another. Here, we need to start by logically thinking about whether one of our concepts could possibly

affect the other. (And no, a logical explanation is not one in which the atrociousness of Nicholas Cage's movies inspires mass drownings.) If we can't think of any causal mechanisms that might conceivably connect these two concepts, we have grounds to conclude that the relationship we observed is spurious. In addition to using logic, another way to rule out chance correlations is to collect more data. Indeed, the "strong" correlations shown on the "Spurious Correlations" website very much depend on the choice of the years of data being shown. If you use a broader date range, you should see these chance correlations disappear.

The other reason that we come across spurious correlations is more serious and difficult to overcome. As we noted earlier, a third concept, the confounder, may be related to our two concepts of interest, driving changes in both. As a result, we may think that shifts in concept A cause shifts in concept B when really a *third* concept lurking in the shadows is causing these changes—for this reason, social scientists sometimes call confounders *lurking variables*. (In **Chapter 12: Experiments** and **Chapter 14: Quantitative Data Analysis**, we'll talk about how we can use controls—control groups or statistical controls, respectively—to address confounders.)

A somewhat related issue is **reverse causality**. In certain cases, we may believe that a change in concept A causes a change in concept B, but the opposite is actually the case (see **Figure 3.3, Diagram 6**). Consider the example we mentioned earlier: is it the case that violent video games cause aggression in teenagers, or do teenagers who already have aggressive tendencies—and who would have them regardless of whether or not they touched a violent video game—just prefer to play these sorts of games? The correlation we observed in the data between playing violent video games and having aggressive tendencies could be evidence of either scenario. To put it another way, both causal stories are possible based on the data we have. We don't have a clear-cut answer to our *why* question about the reasons behind the correlation we've observed.

It is important to remember that a simple correlation between two variables doesn't tell us in which direction the causal arrow points. Just like correlation is not causation, correlation by itself says nothing about which concept is driving the changes we're seeing. As it does with spurious correlations, however, sometimes logic can come to the rescue, telling us whether it's realistic to think that the relationship we're studying goes in one direction rather than another. Here, we're using theory (logic) to guide our inquiry, rather than just randomly identifying correlations. An easy example would be the relationship between gender and your income later in life—since the second concept comes after the first (your birth), it's not possible for the causality in this case to run in any direction other than from gender to income.

Another way to address the problem of reverse causality is to collect data that allows us to decide among different explanatory stories. For example, if we conducted a longitudinal study that followed teenagers over multiple months or years, we could see whether a teenager who started playing violent video games saw their measures of aggression increase afterward. In a similar way, an experiment could tell us in a more convincing fashion whether exposure to violent video games changes people's level of aggression, since we could see the "before" and "after" of that exposure among the study's participants (and, crucially, compare what we find with the same measures for a control group). Note, however, that these alternative approaches have drawbacks, too: longitudinal studies are expensive and time-consuming, and experimenters will find it hard to simulate in a lab setting the actual exposure to video games that children receive in the real world. (We'll discuss the challenges of assessing causality further in **Chapter 12: Experiments**).



Critics of violent video games argue that they make children more aggressive and may have even played a role in the dramatic increase in school shootings over recent decades. The hypothesis that playing a violent video game (say, the first-person shooter game *S.T.A.L.K.E.R.: Call of Pripyat*, pictured here) causes aggression in children is hard to test, however. For one thing, there are numerous potential *confounders*—variables that might lead to both the playing of violent video games and aggression in children (such as a lack of parental attention). Also, the direction of the relationship may go in the opposite direction (*reverse causality*): any correlation between these two variables may be because aggressive kids prefer playing violent video games, rather than because violent video games make kids aggressive. *GSC Game World, via Wikimedia Commons*

So far, we've talked about the danger of reverse causality—believing that changes in one concept cause changes in another when the opposite is actually true. But what if the causal arrow points in *both* directions? In this case, we would say the relationship is **bidirectional** (see **Figure 3.3, Diagram 7**). Bidirectional relationships are quite common in the social sciences: when you're studying complex human beings, there are a lot of moving parts to consider, and social phenomena can frequently be both causes *and* effects. For example, does education cause people to be smarter, or do smart people just get more education? Probably (hopefully, for our jobs as professors), both causal stories are true. This issue, however, makes it exceedingly difficult to assess the "true" impact of going to school on people's measured abilities. Specifically, unless researchers make adjustments, the analysis they conduct will overstate the impact of your education on your smarts, because part of what they are seeing is actually the effect of your innate intelligence on how much education you later receive.

To make matters more complex, we can also come across **feedback loops**, where a change in concept A leads to a change in concept B—which, in turn, loops back to change concept A (see **Figure 3.3, Diagram 8**). A classic example of a feedback loop would be what the sociologist Robert Merton (1948) called "self-fulfilling prophecies." For instance, a person may take out their money from a bank, believing that the bank is on the verge of collapse. Every withdrawal, however, will prompt others to do the same, eventually leading to a collapse even if the bank was initially on firm financial ground. Referring to the diagram can help you visualize the steps in the process: an initial skepticism about the bank (concept A) leads to withdrawals (concept B), which loop back to create more skepticism (concept A), which in turn leads to more withdrawals (concept B). A more colloquial term often used to describe these self-reinforcing relationships is "vicious circles" (sometimes transmuted into "vicious cycles"), with the reverse, a "virtuous circle," used to describe similar chains of positive results.

Social scientists have developed advanced techniques to isolate the causal effect of one concept on another concept, accounting for everything else. Nevertheless, judging whether and how a relationship is causal remains a tricky business, and sociologists need to be cautious about how they describe relationships—and humble about what they can legitimately conclude from their data. At this stage in your research career, just keep these pitfalls in mind as

you judge the validity of your work or other people's work. Use concept maps to make the relationships between the variables you're interested in more tangible and intuitive. Consider each concept in terms of its operational definition as an independent, dependent, moderating, or mediating variable. Imagine how your concepts might relate to your other concepts of interest—and what missing concepts might fill in any gaps or better explain any relationships. Thinking like a researcher requires being able to think clearly and creatively about the many ways a set of concepts can be connected.

Deeper Dive: Idiographic and Nomothetic Explanations



The excuse that "the dog ate my homework" is a tired cliché, but an unfortunate scenario much like this befell Victor as a graduate student when he was working with a professor on a book. In his case, his cat Houdini (not pictured) jumped on his laptop, breaking it and causing him to lose all the revisions he had made to one chapter—the day before the draft was due to his professor. An *idiographic explanation* would consider the many factors that had contributed to that outcome, such as the closeness of the tabletop to the floor, the spryness and girth of the cat, and the decrepitness of the computer. A *nomothetic explanation*, however, would focus on one or a few key variables and look across cases—say, whether having high-jumping cats in one's household is associated with laptop destruction within the larger student population. *Joice Rivas, via Pexels*

Causal explanations can be idiographic or nomothetic. **Idiographic explanations** are those that explain a single situation or event in idiosyncratic detail. For example, suppose you did poorly on an exam. Why did this happen? Well, you forgot that you had an exam on that day. You also arrived late to class because of a traffic jam. Then, midway through the exam, you began panicking and couldn't concentrate. But perhaps most important, the week earlier your dog ate your textbook, and because it was not an open-access textbook that you could access for free on the web, you were unable to study. All these factors played a role in bringing about your poor grade on the exam, and a thorough idiographic explanation would address all of them in order to truly understand why that tragic outcome occurred.

That said, as detailed, accurate, and valid as idiographic explanations like this one can be, they do not necessarily apply to other situations, even ones involving the same person, and are therefore not generalizable. In contrast, **nomothetic explanations** seek to explain a class of situations or events rather than a specific situation or event. For example, we might want to figure out what factors best predict the sorts of students who will do poorly on exams—that is, whether their test grades are most strongly correlated with their family background, their work ethic, their mental health, or other characteristics. Because nomothetic explanations are designed to be generalizable across situations, events, or people, they tend to be less precise, less complete, and less detailed. However, they explain economically, using only a few explanatory variables.

As we've discussed, theories are intended to serve as generalized explanations for phenomena, and they are crucial to the work that sociologists do, given our broader interest in understanding how social life operates for all people at all times. As a result, most sociological research ultimately seeks to arrive at nomothetic explanations. Nevertheless, idiographic explanations can play an important role in research. Especially in qualitative research, we often want to

understand the complex situations that individuals find themselves in—when many factors are at play—rather than reducing their experiences to simple models. Understanding a highly specific context can help sociologists generate creative and surprising theories that can later be considered more broadly.

What Are the Limits of Our Theory?



Army veteran rescues bald eagle dangling upside down from a rope in 75-foot tree by 'mowing down the branches' with his rifle

By Valerie Edwards For Dailymail.com
Updated at 11:35 PM on 2 July 2016



Be aware that the particular context in which you study a phenomenon places limits, or *scope conditions*, on any theory you develop from your data. For example, past studies had described how jobseekers saw networking as essential for getting hired. But those theories were based on research in the United States, a nation that is exceptional in many aspects of its culture. Would the same pattern be seen in Denmark, the Scandinavian country that brought us the concept of *hygge*, a mindset of chill contentment? Sabina Pultz and Ofer Sharone (2020) sought to find out in a qualitative study (described next) based on in-depth interviews with unemployed professionals in both countries. *Jason Galvin and the* Daily Mail*, via Reddit / Stella Rose, via Unsplash*

A final important component of theories is a set of **scope conditions**, also known as *boundary conditions*. All theories are constrained by scope conditions, which tell us where the theory can and cannot be applied. For example, a theory may be bounded, or limited, by culture. Perhaps it is based on data from countries where people tend to be more individualistic in orientation, but it has not been tested in cultures where people have a stronger devotion to the well-being of the collective. Sabina Pultz and Ofer Sharone (2020) conducted a cross-national study of Danish and American unemployed professionals for this very reason. The existing literature emphasized how jobseekers engage in the emotional labor of networking to get hired. But that past work was largely conducted in highly individualistic societies like the United States. In a more egalitarian and collectivist country like Denmark, Pultz and Sharone found, jobseekers tended to be more skeptical about networking, seeing it as potentially exploitative, corrupt, and inauthentic.

Not recognizing the scope conditions of existing studies means that we might take a phenomenon to be widespread when it's very much limited by what sorts of populations and places it has been previously studied in. For example, a theory may have been studied among children, but it is unclear whether it also applies to adults. Or, the observations underpinning a theory may have been done in localities that are completely different from other places and cultures where the same phenomenon occurs, making us question whether the theory is actually relevant there. Indeed, as we

will discuss further in **Chapter 6: Sampling**, the psychological and sociological knowledge that we've gleaned from lab experiments has disproportionately come from so-called **WEIRD societies**—Western, educated, industrialized, rich, and democratic—which raises the perennial question of whether those findings apply more universally.

If we are to properly use or test a given theory, we must be explicit about all of its underlying assumptions and the appropriate boundaries of that theory—where it is valid, and where it is not. When you write up your research, be transparent and exhaustive in describing these scope conditions. When you read other people's work, identify the true boundaries of their theory based on what data they actually have. Otherwise, the theories derived from this research might be applied to contexts they don't fit—leading to erroneous conclusions.

Every study has scope conditions, and being straightforward about the limits of your study's theoretical implications will not signal that your work is poor research—in fact, the opposite is true. We sociologists value honesty and humility about our findings. After all, we can't study a phenomenon in every place on Earth, and yet not doing so means that we must always question whether things will be different in some other setting. (This is not just a problem in the social sciences—yes, the laws of physics hold in every place we've tested them on Earth and outer space, but what about in another solar system, galaxy, or universe?) That said, the fact that a phenomenon has been studied in one context but not anything doesn't necessarily mean it's not applicable elsewhere—the key question is if that context is different in specific ways that would change our findings, as we'll discuss later in this textbook.

Key Takeaways

1. Theories identify certain concepts of interest (in answer to the what question), they determine how those concepts relate to one another (the how question), they explain the reasons behind that relationship (the why question), and they describe the limits of what can be said about that relationship based on the available data (the who, when, and where questions).
2. To test theories, researchers find ways of measuring those concepts (variables), develop hypotheses that relate those variables, and then analyze data that can support or reject those hypotheses.
3. Understanding the relationship between two concepts is complicated by the fact that other concepts may mediate, moderate, or confound that relationship, and the causal arrow connecting these concepts may point in either or both directions.

3.4. Evaluating Theories

Learning Objectives

1. Identify criteria for assessing whether a theory is good or bad.
2. Understand the conditions that have to be in place for a relationship to be causal.

To recap what we've covered so far, using theories in research has many benefits. First, theories can tell us how, why, when, and where a particular phenomenon occurs. They explain the underlying logic and processes (what we've called *mechanisms*) that drive the appearance of that phenomenon. They tell us where and when we might expect to see it, and what outcomes and consequences it has.

Theories also help us make sense of the existing research on a particular topic. A good theory can inspire us to see existing theories in a new light. We can place past studies within a single *theoretical framework* (a collection of theories and related empirical observations). This lets us better see how everything fits together. It gives us insights into the significance and underlying assumptions of past research. And it guides future research, by helping us to identify concepts and relationships that are worthy of further study. As we gather more data, we can reconcile any contradictory findings by expanding or refining existing theories to account for them. We can also develop new theories to bridge gaps between existing ones—even between sociology and other disciplines.



Many scientists labor with the expectation that the theories they work with are getting better and better: with ever-more data-gathering and theorizing, the scientific community's understanding improves, the theory becomes more refined, and scientists' view of reality sharpens. As we'll describe later, this positivist view of science advancing continually toward the "truth" has been challenged by theorists, but it also has been challenged by public misunderstanding and skepticism of science—as seen in the persistence of "flat earth" theories, contrary to all scientific evidence. *Kajetan Sumila, via Unsplash*

That said, it is important for researchers to recognize that the explanations their theories provide may be good or poor. There is nothing sacrosanct about any theory, and theories should not be accepted just because they were proposed by someone well-known and respected. Indeed, science makes progress when scientists see that a theory falls short in a particular area, and then refine or rework the theory so that it better describes, explains, and predicts reality. In the course of this work, poorer theories are eventually replaced by better theories with higher explanatory power. The essential challenge for researchers is to build better and more comprehensive theories that can explain a phenomenon more accurately and comprehensively than prior ones.

We should also keep in mind that every theory is incomplete—they are by definition simplifications of reality. Especially if we want them to help guide our decisions, theories have to reduce a phenomenon to a limited set of concepts and relationships. As social scientists, we're always balancing the need between making our theories simple enough that they are understandable and actionable, and making them sophisticated enough that they accurately depict the world around us. Furthermore, scholars naturally tend to gravitate toward certain theoretical perspectives (more about that later), which may impose blinders and limit their sociological imagination. Seeing the world from a particular vantage point, they may miss out on important concepts that are not adequately addressed by their pet theory. We should keep in mind that theories—even our favorite ones—will rarely provide a full or utterly convincing explanation for a given phenomenon. Social life is too complex for that to be the case.

So, how can we evaluate theories? Here are some common criteria for judging whether a given theory is "good" or "poor":

1. **Logical consistency:** Are the theory's concepts, hypotheses, and scope conditions consistent with each other? Any contradictions might mean the underlying logic of the theory is flawed. For example, suppose a theory assumes that people are *rational actors*—making logical decisions based on the costs and benefits of any actions. If we are studying situations in which people tend to make nonrational decisions—based on emotion rather than reason, for instance—the theory will probably not be a convincing explanation of people's behavior within these scenarios.
2. **Explanatory power:** How much does a given theory explain (or predict) reality? Good theories obviously explain the target phenomenon better than rival theories. For example, if one theory argues that a college degree pays off because educated people are smarter workers, and another theory argues that employers just pay people more who have that credential, then one of these theories should be able to account for more of the pay bump that college graduates receive than the other theory.
3. **Falsifiability:** The British philosopher Karl Popper once said that for theories to be valid, they must be **falsifiable**.

In other words, it must be possible to prove that a hypothesis generated from the theory is untrue based on empirical observations. Tautological statements, such as "a day with high temperatures is a hot day," are circular in their reasoning; they are not empirically testable because a hot day is defined (and measured) as a day with high temperatures. As a result, such statements cannot serve as hypotheses, and the underlying theory is not falsifiable. Furthermore, whenever we evaluate a theory, there should be rival (alternative) explanations for the phenomenon in question. When we were discussing the payoff to a college degree earlier, we noted such competing explanations: one being that employers pay college graduates more for their actual skills, and the other being that they pay more for a "signal" of that skill—the college degree. We can decide between these two theories—and falsify one of them—by, for instance, seeing how much people with the same level of skill but differing credentials are paid. (As it turns out, there is a large bump in pay between people with almost enough credits to graduate college and those who do graduate—which suggests that it's really the *signal* that employers focus on.) Saying that a theory is falsifiable, however, is not the same as saying that a theory *should* be falsified—you have better things to spend your time on than proposing a theory that you think will turn out to be false. Rather, the important thing is that your theory can possibly be proven false. Otherwise, there's no point in conducting research to test that theory—your study already has a big thumb on the scale in its favor.

4. **Parsimony: Parsimony** refers to how economical a theory is—how much of a particular phenomenon it can explain, with how few variables. The fourteenth-century English logician William of Ockham (also written as "Occam") developed a principle for making theories parsimonious, which is called "Occam's razor" in his honor. According to this principle, of all the competing explanations that explain the observed evidence, the simplest theory—the one that uses the fewest variables or makes the fewest assumptions—is the best. After all, our ability to explain a complex social phenomenon can always be enhanced by adding more and more concepts. Such an approach, however, defeats the purpose of having theories, which, as we noted earlier, are meant to be simplified explanations of reality. Parsimonious theories can be more easily generalized to other contexts, settings, and populations, and we should favor them if their explanatory power is the same as that of more complex theories.

The second item, *explanatory power*, merits further discussion, given how crucial it is to identify and understand cause-and-effect relationships in the social sciences. Here are four criteria for a good causal theory:

1. **Correlation:** A change in one variable is associated with a change in the other. (This is the bare minimum requirement for causality—remember, "correlation is not causation"!)
2. **Temporal precedence:** The cause occurs before the effect. Here, we're interested in the *sequence* of events: the change in the independent variable has to take place before the change in the dependent variable, or the former cannot logically be considered a cause of the latter.
3. **Rejection of alternative hypotheses:** After testing whether other variables explain the phenomenon of interest, we conclude that these other possibilities are less convincing.
4. **Causal mechanisms:** We can describe the specific process by which changes in the independent variable lead to changes in the dependent variable.



William of Ockham (c. 1287–1347), an English friar and philosopher, argued that theories should be efficiently reasoned, with any unnecessary aspects of a theory—any assumptions or causal factors not strictly necessary to explain things—shaven away. This maxim to be parsimonious in one's theorizing was named "Occam's razor" in his honor. *AldrianMimi, via Wikimedia Commons*

Some researchers might just stop at the first step—they care about correlations, and not about whether a change in one variable truly *causes* a change in another. Predicting events or behaviors using a set of variables can be useful in itself, even if you can't necessarily explain why those events or behaviors are taking place. For example, market analysts routinely predict fluctuations in the stock market based on market announcements, earnings reports of major companies, and new data from the U.S. Federal Reserve or similar government agencies. Past research has identified correlations between those regularly observed variables (what are called **indicators**) and the overall performance of the market. Based on those past associations, analysts can plug in the current values of the relevant variables into their statistical models and predict what the market will look like at some later point. The same goes for the results of future elections: social scientists often use a mix of political polls, economic indicators, and other measures of voter mood to predict who will win elections—sometimes far into the future. The relationships here are not necessarily causal, but prediction requires us only to measure *correlation*. Indeed, the models we use in these cases may be rather stripped down—as we have described, models are always useful simplifications—with just some "inputs" (the values of our independent variables) that the model somehow converts into "outputs" (the predicted values of our dependent variable).

As social scientists, however, we usually want to do more than just observe associations between variables. We want to explain them. We want to understand how exactly a change in one variable leads, logically, to a change in the other. Scientists want to peer inside the "black box" that connects the input and the output—the independent variable and the dependent variable—and comprehend what's going on there. That's what scientific knowledge is about.

Key Takeaways

1. Good theories are logically consistent, falsifiable, and parsimonious explanations of the relationships between concepts.
2. To show that a relationship between two concepts is causal, the researcher must go much further than simply identifying a correlation. The change in the independent variable must precede the change in the dependent variable, and any alternative explanations must be ruled out. A good causal explanation will also have concrete causal mechanisms in mind.

3.5. Major Sociological Theories and Paradigms

Learning Objective

Identify and describe the dominant theoretical perspectives and paradigms in the social sciences.

Social theories provide a way of looking at the world and understanding human interaction. In the previous sections, we noted that theories are not just collections of facts; they need to put forward some sort of explanatory story. At the same time, theories operate at many different levels, not just in terms of their units of analysis, but also in how grand or sweeping—or restricted or specific—an explanation they provide.

Most theories (especially the ones that you, as a novice researcher, would tackle for your first study) are narrow in focus. They might aim to provide a compelling explanation for one particular phenomenon, in one specific context. Moving up the ladder of ambition and complexity, however, we arrive at cohesive collections of theories—what we might think of as "major" or "overarching" theories, or *theoretical traditions* or *theoretical perspectives*—that aim to explain the very existence and persistence of society as we know it. Introductory sociology courses typically teach students about the "big three" sociological theories: structural functionalism, conflict theory, and symbolic interactionism. As you might recall, **structural functionalism** focuses on the interrelations between various parts of society and how each part works with the others to make society function in the way that it does—much like parts of the body work together to help an organism to thrive. **Conflict theory** is interested in questions of power. Scholars in this tradition examine who wins and who loses based on the way that society is organized. **Symbolic interactionism** studies how meaning is created and negotiated through social interactions. Sociologists who use this approach focus on **symbols**—that is, the ways that people communicate their thoughts and feelings.

Just as researchers can examine the same topic from different levels of inquiry, so, too, can they investigate the same topic from different theoretical perspectives. In this case, even their research questions could be the same, but the way they make sense of whatever phenomenon it is they are investigating will be shaped in large part by their choice of a theoretical perspective and the assumptions that underlie it.

Table 3.1 summarizes the major points of focus for each of the three major theories and outlines how a researcher might approach the study of the same topic—in this case, the study of sports—from each of the three perspectives.

Table 3.1. Major Sociological Theories and the Study of Sports

Theoretical Perspective	**Focuses on . . .**	**A Study of Sports Might Examine . . .**
Structural functionalism	Interrelations between parts of society, and how those parts work in concert	The consequences—positive and negative, intended and unintended—of professional sport leagues
Conflict theory	Who wins and who loses based on the way that society is organized	Issues of power in sports, such as differences in participation and access to training
Symbolic interactionism	How meaning is created and negotiated through interactions	How the rules of sports are constructed, taught, and learned



The major theoretical traditions in sociology tend to emphasize different aspects of the same social phenomenon. For instance, sociologists applying a structural functionalist perspective to their study of sporting events might delve more deeply into how athletes promote teamwork and cultivate camaraderie. Researchers attuned to symbolic interactionism would pay particular attention to the ways that individuals and teams communicate on the playing field, such as through the use of signals, mannerisms, and language that convey meanings specific to the context. *Anastasia Shuraeva, via Pexels / Nathan Shively, via Unsplash*

Within each area of specialization in sociology, there are many other theories that aim to explain more specific types of interactions. For example, within the sociological study of sexual harassment, different theories suggest different explanations for why harassment occurs. One theory, first developed by criminologists, is called *routine activities theory*. It argues that sexual harassment is most likely to occur when a workplace lacks unified groups, and when potentially vulnerable targets and motivated offenders are both present (De Coster, Estes, and Mueller 1999). Other theories of sexual harassment, called *relational theories*, suggest that a person's relationships, such as their marriages or friendships, are the key to understanding why and how workplace sexual harassment occurs, and how people will respond to it when it does occur (Morgan 1999). Relational theories focus on the power that different social relationships provide. For example, married people who have supportive partners at home might be more likely than those who lack

support at home to report sexual harassment when it occurs. Finally, feminist theories of sexual harassment argue that the way our current gender system is organized—giving the most power to those who are the most masculine—best explains why and how workplace sexual harassment occurs (MacKinnon 1979).

As is true for the study of any topic, which theory you apply to examine sexual harassment will shape which research questions you ask. It will also influence the sorts of explanations you provide for why harassment occurs. Given how complex our social realities are, all of these theories may be true at the same time, though some may work better in certain contexts. Research can help us weigh the impact of these different factors.

Paradigms versus Theories: Why and How



The American philosopher Thomas Kuhn (1922–1996) advanced an influential theory that the sciences make progress in fits and starts. At any given time, a master theory—a *paradigm*—will dominate a particular scientific field. Gradually, scientific evidence will accumulate to challenge that paradigm. While resisted at first, the contrary theories will eventually upend key aspects of the paradigm, leading to what Kuhn called a *paradigm shift*—a movement toward a new scientific consensus that reconciles older findings within the latest paradigm. *Davi.trip, via Wikimedia Common*

Paradigms stand at an even higher level of abstraction than the theoretical perspectives we were just discussing. The terms "paradigm" and "theory" are often used interchangeably in social science, but we follow other researchers in distinguishing between the two ideas, which can help us to understand the connections between research methods and social scientific ways of thinking. More encompassing than mere theories, paradigms are ways of viewing the world and understanding the human experience. They frame what we know, what we can know, and how we can know it.

The term "paradigm" was popularized by the philosopher Thomas Kuhn. In his book *The Structure of Scientific Revolutions* (1962), Kuhn examined the history of the natural sciences to identify patterns of activities that determine the progress of science. As mental models or frames of references, paradigms organize our reasoning and observations, Kuhn argued. As belief systems, they contain underlying, and often unacknowledged, assumptions. In these ways, paradigms fundamentally shape the ways that scientists design and conduct research. At any given moment, a field of science may be dominated by a particular paradigm; as more research accumulates that challenges that paradigm, however, the field may undergo what Kuhn called a *paradigm shift*, during which the field's scientists gradually adopt a new way of understanding reality that better fits the available evidence. (Most famously, this paradigm shift occurred when Isaac Newton's classical theories of physics were replaced by Albert Einstein's theories of relativity.)

Each individual's viewpoint is shaped by many paradigms—from moral and political ones, to social scientific ones. These paradigms are like "colored glasses" that govern how we view the world and how we think about what we see in the world. Like colored glasses, we become used to them over time and may forget they're even there. Indeed, it can be difficult to fully grasp the idea that our way of looking at the world reflects a particular paradigm and set of accompanying assumptions. As individuals, we are very much steeped in our own personal and everyday perspective. For example, consider people's views on abortion. To some, abortion is a medical procedure that should be undertaken at the discretion of each individual who might experience an unwanted pregnancy. To others, abortion is murder, and members of society should collectively have the right to decide when, if at all, abortion should be undertaken. If you have an opinion about this controversial topic, you might be pretty certain that you are "right." The person who sits next to you in class may have a very different opinion and yet be equally confident about the truth of their perspective.



Adopting a new theoretical perspective can be incredibly liberating, revealing your taken-for-granted assumptions and making you wonder why you've never seen things in this way before. At the same time, there are dangers in viewing the world through tinted glasses of any hue: our social reality is too complex to reduce to one set of theories, however compelling they may be. If you are too fixated on a particular theoretical point of view, you might unwisely see the solution to every problem in those terms: as the adage goes, if all you have is a hammer, everything looks like a nail. *Reymon Reymon, via Pexels*

Which of you is correct? You are each operating under a set of assumptions about the way the world does—or at least should—work. Perhaps your assumptions come from your particular political perspective, which guides your view on a variety of social issues. Or perhaps your assumptions are based on what you learned from your parents or the community (religious or otherwise) you were raised in. In either case, a paradigm shapes your understanding of the issue and therefore your stance on it.

You can think of theories as being nested within larger paradigms, providing different sorts of answers to our questions. As we noted earlier, theories help us answer the *why* questions we often have about the patterns we observe in social life. But even before we arrive at that theory, the paradigm we are operating under may point us in a particular direction with respect to our *why* questions. For better or worse, it may constrain our thinking and reasoning about what we observe. For instance, people may have very different opinions about how to solve social problems depending on what they believe the proper role of government should be in people's lives. On the political right, conservatives and libertarians may believe that lowering taxes is the best way to stimulate a stagnant economy because it increases people's disposable income and spending, which in turn expands business output and employment. On the political left, liberals and progressives may believe that governments should invest more directly in job-creation programs, such as public works and infrastructure projects, which will increase employment and people's ability to consume, spurring greater economic output. Likewise, Western societies tend to emphasize individual rights, such as people's right to privacy, freedom of speech, and right to bear arms. In contrast, Asian societies often choose to balance the rights of individuals against the rights of families, organizations, and the government in ways that favor more communal and less individualistic policies. Such differences in perspective often lead Westerners to criticize Asian governments for being autocratic, while Asians criticize Western societies for being greedy and selfish. You might imagine how researchers from each of these backgrounds might take very different approaches to studying the same social policies or social norms—or how they might find entirely different topics worth studying in the first place.

As we've noted, paradigms are often hard to recognize because they are implicit and taken for granted. But identifying

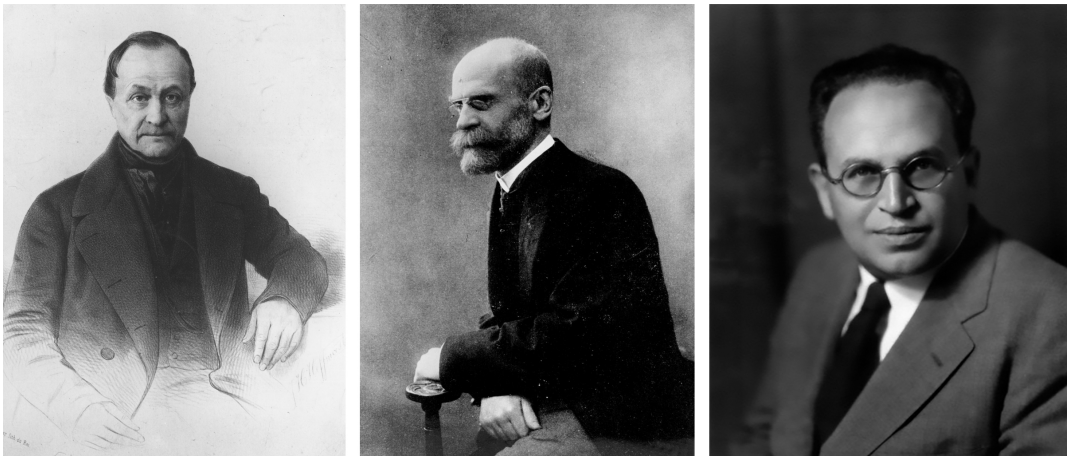
which paradigm a particular person is operating under is key to making sense of and reconciling differences in people' perceptions of the same social phenomenon. For instance, if you ask the public what is the best way to improve the public education system, liberals and progressives might prefer paying teachers higher salaries, while conservatives and libertarians might prefer expanding the use of school vouchers—allowing parents to use tax dollars to send their children to private schools. What accounts for these differences in policy preferences are distinct assumptions that people on the left and right bring to their analysis of the educational system. On the right, people tend to place more faith in competitive markets: under a voucher system, the logic goes, schools will compete for education dollars, and that competition will create pressures for quality improvements. On the left, people prioritize treating workers well, with the idea that higher wages will create incentives for better teaching.

Paradigms in Social Science

Let's look at four of the most common social scientific paradigms that might guide you as you begin to think about conducting research (**Table 3.2** summarizes each of them). Consciously or unconsciously, these paradigms constrain the concepts that researchers attempt to measure, their observations of a particular phenomenon, and their subsequent interpretations of that phenomenon.

The first paradigm we'll consider, called **positivism**, is probably the framework that comes to mind for many of you when you think of science. Positivism is guided by the principles of objectivity, knowability, and deductive logic. (Deductive logic is discussed in more detail in **Chapter 4: Research Questions**.) The positivist framework operates from the assumption that society can and should be studied empirically and scientifically. It holds that science or knowledge creation should be restricted to what can be observed and measured.

The French philosopher Auguste Comte, the person who coined the term *sociology*, argued that the discipline should be a positivist science (Ritzer and Stepnisky 2017). Blending together notions of rationalism and empiricism, Comte suggested that theory and observations have a circular dependence on each other—that is, they build on each other in ways similar to the feedback loops we mentioned earlier. Specifically, while theories may be created through reasoning, they are authentic only if they can be verified through observations, Comte argued. This emphasis on verification pushed the social scientists of the day to separate themselves from the purely theoretical fields of philosophy and metaphysics and embrace the "scientific method" as the primary means of validating scientific claims.



The French philosopher Auguste Comte (1798–1857) promoted the new field of "sociology"—a term he invented—as a reason-based science that would only hold truths to be evident when they were backed up by empirical observations. The sociologists Émile Durkheim (1858–1917) and Paul Lazarsfeld (1901–1976) followed in this tradition, which eventually led to the field being dominated by quantitative researchers who drew on national survey data and the new science of probability sampling (described in **Chapter 6: Sampling**). *Johan Hendrik Hoffmeister* / *Wikimedia Commons* / *Bardwell Press*

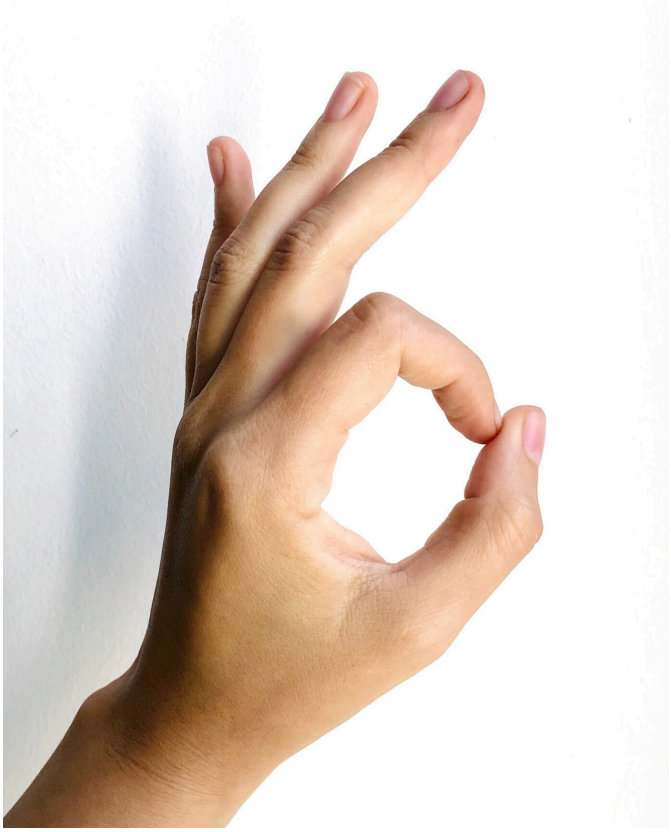
Key scholars who built on Comte's ideas about positivism included Émile Durkheim, who developed *sociological positivism* (positivism as a foundation for social research) and Ludwig Wittgenstein, who developed *logical positivism* (positivism as a foundation for philosophy). In the field of sociology, positivism in sociology often took the form of large-scale survey research. For example, the American sociologist Paul Lazarsfeld pioneered the creation of national surveys, along with new statistical techniques to analyze the quantitative data that these surveys produced.

As we noted, positivism relies on theories that can be directly tested. It was originally an attempt to separate scientific inquiry from religion, whose beliefs could not be objectively tested. But critics charged it went too far in the opposite direction, leading to a blind faith in data and a rejection of any attempts to reason beyond observable facts. In the early twentieth century, positivism came under widespread attack. While positivism called for researchers to abandon their biases and values in a quest for objective, empirical, and knowable truth, critics argued that such a value-free science was impossible. In fact, they accused positivists of being "atheoretical" (a cardinal sin among sociologists)—that is, not

being explicit about the theoretical perspectives they brought to their study of social phenomena, and failing to consider the merits and weaknesses of different theories in explaining their data. Even prominent positivists like Lazarsfeld came to acknowledge potential problems with the approach—for example, how difficult it was to root out biases in the ways that scientists observe the world.

Frustrations with the strictly empirical nature of positivist philosophy led to the development of **postpositivism** (or *postempiricism*) during the mid- to late twentieth century. The British philosopher Karl Popper suggested that human knowledge is not based on unchallengeable, rock-solid foundations. Instead, it amounts to a set of tentative conjectures that can never be proven conclusively, but only disproven. Unlike positivism, postpositivism concludes that it is ultimately impossible to verify the truth, but like positivism, it seeks to reject false beliefs through empirical observation. It argues that one can make reasonable inferences about a phenomenon by combining empirical observations with logical reasoning—in this way, moving eventually toward the objective truth that positivism promises. The postpositivist camp has further fragmented into *subjectivists*, who view the world as a subjective construction of our subjective minds rather than as an objective reality, and *critical realists*, who believe that an external reality exists that is independent of a person's thinking—though we can never know this reality with any degree of certainty.

The critiques of positivism—which fall under the large umbrella of **antipositivism**—have taken many forms. Some common themes can be emphasized, however. While positivism is typically associated with deductive and quantitative research methods, such as experiments and surveys, antipositivism often favors studying social life through interpretive means. Antipositivists are more likely to employ qualitative methods, such as in-depth interviewing and ethnographic observation, to plumb the meaning and purpose that individuals attach to their personal actions—information that is disregarded, antipositivists say, when quantitative methods alone are used.



Social constructionism, a paradigm that challenges positivist conceptions of social science, argues that notions of truth are socially constructed. For example, social constructionists might point out that the wide variety of symbols we use to communicate on a daily basis—including hand signals like the "OK" sign pictured here—only have the meanings that specific cultures impart to them. Our very understandings of reality are constructed from collective agreements, and those agreements are constantly evolving—as might be seen, for instance, in controversies over whether the "OK" hand gesture is a sign used by white nationalists (Allyn 2019). *Elena Rabkina, via Unsplash*

One dominant antipositivist paradigm in sociology is **social constructionism** (Berger and Luckmann [1966] 1991). While positivists seek "the truth," the social constructionist framework argues that "truth" is a varying, socially constructed, and ever-changing notion. This is because reality does not simply exist, waiting for us to "discover" it. According to social constructionism, we create reality ourselves through our interactions and our interpretations of those interactions. Key to this paradigm is the idea that *social* context and interaction frame and shape these interpretations. While individuals may construct their own realities, groups—from ones as small as married couples, to ones as large as nations—often agree on notions of what is true and what "is." For example, research in this paradigm has examined how the meanings of different hand gestures vary across different regions of the world, underscoring that meanings are constructed socially and collectively. Think about what it means to you when you see a person raise their middle finger. In other societies, it is a different gesture, the thumbs-up, that raises eyebrows. The meanings of these two hand signals are not necessarily shared across cultures, and the same goes, social constructionists say, for all the other symbols, expressions, and concepts that we may believe to be universal.

A third paradigm is the critical paradigm (often called **critical theory**). At its core, critical theory is focused on power, inequality, and social change. Critical theory attempts to analyze the oppositions, conflicts, and contradictions in contemporary society, critique the alienating and oppressive conditions of the status quo, and eliminate the causes of that alienation and oppression. Although some rather diverse perspectives are included here, critical theory is rooted in ideas developed by early social theorists, such as Max Horkheimer (1972), and later works developed by feminist scholars, such as Nancy Fraser (1989).





Social theorists like Max Horkheimer (1895–1973) and Nancy Fraser developed a diverse set of perspectives in the late twentieth century that came to be known as *critical theory*, which examines power structures within society with the goal of resisting and overturning them. *Barbara Niggl Radloff and Bunnyfrosch, via Wikimedia Commons*

Like social constructionism, critical theory rejects the view that social science is objective or value-free. Unlike social constructionism, it tends to operate more from the perspective that scientific investigation should be conducted with the express goal of social change in mind. The German philosopher Karl Marx—considered a founding figure of sociology—summed up this viewpoint: "Philosophers have only interpreted the world, in various ways; the point is to change it" (1992:423). Marx critiqued capitalistic societies for their huge divides between the rich and poor and argued that workers needed to overthrow the prevailing economic system. True to Marx's intent, his philosophy of Marxism inspired social revolutions in countries such as Germany, Italy, Russia, and China, although it also led to the rise of despotic regimes and inequalities of other kinds—a reminder of the ways that a value-driven science can also be dangerous.

Finally, **postmodernism** is a paradigm that challenges almost every way of knowing that many social scientists take for granted (Best and Kellner 1991). While social constructionists might argue that truth is in the eye of the beholder (or in the eye of the group that agrees on it), postmodernists roundly reject the notion of any sort of knowable truth—even to that limited degree. We can never really discern the truth, they would emphasize, because in the studying and reporting of others' truths, the researcher stamps their own truth on the investigation. Likewise, while the critical paradigm might argue that power, inequality, and change shape reality and truth, a postmodernist might ask, in turn: whose power, whose inequality, whose change, whose reality, and whose truth? As you might imagine, the postmodernist paradigm poses quite a challenge for social scientific researchers. How does one study something that may or may not be real, or that is real only in your current and unique experience of it?

This does not necessarily mean that social science is fundamentally corrupt, but it is always important to keep in mind that human researchers can never claim to be entirely value-free. As social constructionists and postmodernists emphasize, bias is always a part of research. Our job as researchers is to recognize and address our biases during the course of our research so that they don't overwhelm the value of our work. As with other competing theories, the complex nature of social phenomenon means that all of these paradigms may hold truth. Sociologists who want to arrive

at the fullest possible understanding of the problem at hand could find themselves drawing from multiple paradigms to explain it.

Table 3.2. Social Scientific Paradigms

Paradigm	Emphasis	Assumption
Positivism	Objectivity, knowability, and deductive logic	Society can and should be studied empirically and scientifically.
Social constructionism	Truth as varying, socially constructed, and ever-changing	Reality is created collectively, and that social context and interaction frame our realities.
Critical theory	Power, inequality, and social change	Social science can never be truly value-free and should be conducted with the express goal of social change in mind.
Postmodernism	Inherent problems with previous scientific paradigms	Truth in any form may or may not be knowable.

Key Takeaways

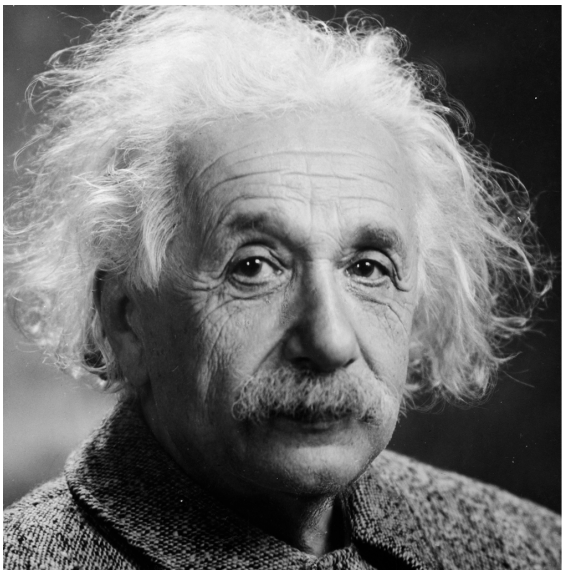
1. Sociologists use theory to help frame their research questions and to help them make sense of the answers to those questions.
2. Some well-established sociological theories (often called theoretical perspectives) are rather sweeping in their coverage and attempt to explain, broadly, how and why societies are organized in particular ways.
3. As even larger "umbrella" theories, scientific paradigms shape our everyday view of the world (and our approach as researchers) in subtle ways.

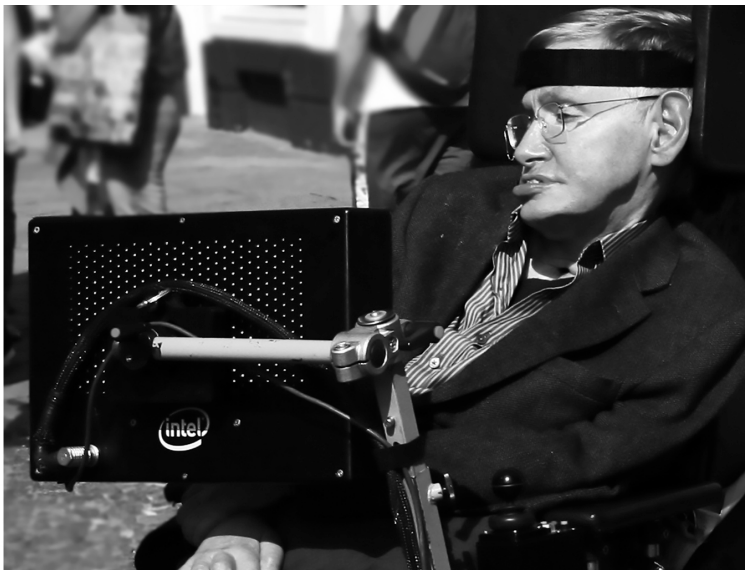
3.6. Linking Theory and Data

Learning Objective

Describe the roles that theory and data play in conducting social scientific research.

As we wrap up this chapter, let's review the ways that data and theories are relevant to sociological research methods. We sociologists employ diverse theories, levels of analysis, and analytical procedures in our work. Understanding the strengths and weaknesses of each of these approaches is crucial not only for successfully conducting research, but also for responsibly reading and evaluating the work of others.





At times, the theorizing that social scientists do might seem out of touch, an activity for academics in the ivory tower who are entirely detached from the real world. This is a fair criticism, but remember that the transformative thinkers in any scientific field—above all, household names like Albert Einstein (1879–1955) and Stephen Hawking (1942–2018)—have always been master theoreticians who used data to make bold claims about the nature of reality. *Orren Jack Turner and Doug Wheller, via Wikimedia Commons*

First, our choice of a level of analysis and a particular type of data will shape what we ask about the social world and how we ask it. A micro-level study of gangs will look much different than a macro-level study of gangs. In some cases, you could apply multiple levels of analysis to your investigation, but doing so isn't always practical. Therefore, it's crucial to understand the different levels of analysis, be aware of which level you happen to be employing, and recognize the inevitable advantages and disadvantages of that choice. Your theoretical perspective will also shape your study. The theory you invoke will influence not only the way a question about a topic is asked, but also which topic gets investigated in the first place. Likewise, the possible answers you find to the questions you pose will be constrained by your choice of an overarching paradigm.

Second, it is important to recognize the fact that good research requires you to be proficient in *both* theory and methodology. Methodological skills—which we will describe throughout the second half of this book—can be boiled down to knowing how best to measure and evaluate data. Social sciences make progress as individual scientists develop more sophisticated ways of counting and describing the world, and you as a researcher can make important contributions by being creative in this regard. In contrast, being "skilled" at theory means that you know the real significance of whatever empirical findings you uncover in your research. You as a theorist will advance social science by developing concepts and identifying patterns that explain the world more fully or accurately.

Having deficits in either your methodological or theoretical skills is a major problem. At times, students with brilliant theoretical minds stumble as sociologists because they don't know or appreciate the nuts and bolts of research methods. They create studies with obvious flaws in logic, or they fail (because of ignorance or laziness) to gather the data they need to confirm or disconfirm their pet theories. Generally speaking, though, theory tends to be considerably harder to master than methodology, requiring years of reading, observation, and reflection. Provocative theories can also revolutionize the field in ways that a competently done, but narrowly theorized, study cannot. The greatest scientists in history—from the earliest astronomers, chemists, and physicists, to the major social scientists of modern times—are usually held in awe for the theories they postulated that transformed the course of science. Strong methodological skills are needed to be an ordinary researcher, but strong theoretical skills are needed to be an extraordinary researcher.

Key Takeaways

1. What our research will end up saying depends on our theoretical and empirical choices. The theory being invoked, and the paradigm from which a researcher frames his or her work, can shape not only the questions asked but also the answers discovered. Likewise, different levels of analysis lead to different points of focus on any given topic.
2. A good researcher needs to cultivate two sets of skills: theoretical and methodological. Mastering theory is particularly difficult, but being able to apply a unique and creative theoretical perspective to compelling data is the best way to ensure that your research has wide influence.

3.7. References

Allyn, Bobby. 2019. "The 'OK' Hand Gesture Is Now Listed as a Symbol of Hate." NPR, September 26.

Bem, Sandra L. 1974. "The Measurement of Psychological Androgyny." *Journal of Consulting and Clinical Psychology* 42:155–62. doi: 10.1037/h0036215.

Berger, Peter L., and Thomas Luckmann. [1966] 1991. *The Social Construction of Reality: A Treatise in the Sociology of Knowledge*. New York: Penguin.

Best, Steven, and Douglas Kellner. 1991. *Postmodern Theory*. New York: Guilford.

Bhattacherjee, Anol. 2012. *Social Science Research: Principles, Methods, and Practices*. Athens, GA: Global Text Project.

Brady, David, and Amie Bostic. 2015. "Paradoxes of Social Policy: Welfare Transfers, Relative Poverty, and Redistribution Preferences." *American Sociological Review* 80(2):268–98. doi: 10.1177/0003122415573049.

De Coster, Stacy, Sarah Beth Estes, and Charles W. Mueller. 1999. "Routine Activities and Sexual Harassment in the Workplace." *Work and Occupations* 26(1):21–49. doi: 10.1177/0730888499026001003.

Deleuze, Gilles, and Felix Guattari. 2009. *Anti-Oedipus: Capitalism and Schizophrenia*. New York: Penguin Classics.

Donato, Katharine M., and Gabriela León-Pérez. 2017. "Educación, Género y Migración de Colombia y México a Estados Unidos." Pp. 179–202 in *Migración Internacional Patrones y Determinantes: Estudios Comparados Colombia-America Latina-Proyecto LAMP*, edited by M. G. Roa Martínez. Cali, Colombia: Universidad del Valle.

Dordick, Gwendolyn. 1997. *Something Left to Lose: Personal Relations and Survival among New York's Homeless*. Philadelphia: Temple University Press.

Durkheim, Émile. 1897. *Suicide: A Study in Sociology*. Florence: Taylor & Francis.

Frank, David John, Bayliss J. Camp, and Steven A. Boutcher. 2010. "Worldwide Trends in the Criminal Regulation of Sex, 1945 to 2005." *American Sociological Review* 75(6):867–93. doi: 10.1177/0003122410388493.

Fraser, Nancy. 1989. *Unruly Practices: Power, Discourse, and Gender in Contemporary Social Theory*. Minneapolis: University of Minnesota Press.

Freedman, Neal D., Yikyung Park, Christian C. Abnet, Albert R. Hollenbeck, and Rashmi Sinha. 2012. "Association of Coffee Drinking with Total and Cause-Specific Mortality." *New England Journal of Medicine* 366(20):1891–1904. doi: 10.1056/NEJMoa1112010.

Hochschild, Arlie Russell. 1979. "Emotion Work, Feeling Rules, and Social Structure." *American Journal of Sociology* 85(3):551–75.

Horkheimer, Max. 1972. *Critical Theory: Selected Essays*. New York: Herder & Herder.

Khan, Shamus Rahman. 2011. *Privilege: The Making of an Adolescent Elite at St. Paul's School*. Princeton, NJ: Princeton University Press.

Kimmel, Michael S., and Amy Aronson. 2004. *Men and Masculinities: A Social, Cultural, and Historical Encyclopedia*. Santa Barbara, CA: ABC-CLIO.

Kuhn, Thomas S. 1962. *The Structure of Scientific Revolutions*. 2nd ed. Chicago: University of Chicago Press.

Lamont, Michèle. 1994. *Money, Morals, and Manners: The Culture of the French and the American Upper-Middle Class*. Chicago: University of Chicago Press.

Lareau, Annette. 2011. *Unequal Childhoods: Class, Race, and Family Life*. 2nd ed. Berkeley: University of California Press.

Leidner, Robin. 1993. *Fast Food, Fast Talk: Service Work and the Routinization of Everyday Life*. Berkeley: University of California Press.

MacKinnon, Catharine A. 1979. *Sexual Harassment of Working Women*. New Haven, CT: Yale University Press.

Marks, Stephen, Ted Huston, Elizabeth Johnson, and Shelley Macdermid. 2001. "Role Balance among White Married Couples." *Journal of Marriage and Family* 63:1083–98. doi: 10.1111/j.1741-3737.2001.01083.x.

Marks, Stephen R., and Shelley M. MacDermid. 1996. "Multiple Roles and the Self: A Theory of Role Balance." *Journal of Marriage and Family* 58(2):417–32. doi: 10.2307/353506.

Marx, Karl. 1992. *Early Writings*. London: Penguin.

Merton, Robert K. 1948. "The Self-Fulfilling Prophecy." *Antioch Review* 8(2):193–210. doi: 10.2307/4609267.

Messner, Michael. 2009. *It's All for the Kids: Gender, Families, and Youth Sports*. Berkeley: University of California Press.

Morgan, Phoebe A. 1999. "Risking Relationships: Understanding the Litigation Choices of Sexually Harassed Women." *Law and Society Review* 33(1):67–92. doi: 10.2307/3115096.

Osuji, Chinyere K. 2019. *Boundaries of Love: Interracial Marriage and the Meaning of Race*. New York: NYU Press.

Pultz, Sabina, and Ofer Sharone. 2020. "The Intimate Dance of Networking: A Comparative Study of the Emotional Labor of Young American and Danish Jobseekers." Pp. 33–58 in *Professional Work*. Vol. 34, *Research on the Sociology of Work*, edited by E. H. Gorman and S. P. Vallas. Bingley, UK: Emerald.

Putnam, Robert D. 2001. *Bowling Alone: The Collapse and Revival of American Community*. London: Touchstone Books by Simon & Schuster.

Ritzer, George, and Jeffrey N. Stepnisky. 2017. *Classical Sociological Theory*. 7th ed. Los Angeles: SAGE.

Ruppanner, Leah E. 2010. "Cross-National Reports of Housework: An Investigation of the Gender Empowerment Measure." *Social Science Research* 39(6):963–75. doi: 10.1016/j.ssresearch.2010.04.002.

Sampson, Robert J. 2013. *Great American City: Chicago and the Enduring Neighborhood Effect*. Chicago, IL: University of Chicago Press.

Spicer, Jason, and Christa R. Lee-Chuvala. 2021. "Ownership and Mission Drift in Alternative Enterprises: The Case of a Social Banking Network." Pp. 257–91 in *Organizational Imaginaries: Tempering Capitalism and Tending to Communities through Cooperatives and Collectivist Democracy*. Vol. 72, *Research in the Sociology of Organizations*, edited by K. K. Chen and V. T. Chen. Bingley, UK: Emerald.

Vigen, Tyler. 2015. *Spurious Correlations*. New York: Hachette.

Whetten, David A. 1989. "What Constitutes a Theoretical Contribution?" *Academy of Management Review* 14(4):490–95. doi: 10.5465/amr.1989.4308371.

4. RESEARCH QUESTIONS



Designing a new study can seem daunting, especially after you've read studies by professional sociologists. With all of the details to consider, and all of the things that can go wrong, how can you put together a feasible and potentially useful study? It's important to remember that the design of a study evolves over time, and you typically won't get it right in the first draft. Doing research is a marathon, not a sprint, and you want to focus on a single step at a time, knowing that as long as you're making incremental progress, you'll eventually reach your destination. We also want to emphasize how the planning process for a scientific study—what we call **research design**—tends to be long and involved. Social scientists are expected to do a lot of background reading and thinking about their research before they start collecting or analyzing data to any substantial degree. Unlike journalists, social scientists don't usually work on tight daily deadlines, and they take the time to ensure that their work truly builds on existing knowledge.

You will start the research design process by developing an interesting **research question**, the question you are trying to answer by collecting and analyzing data for your study. This chapter will focus on strategies for how to go about generating a good research question. We want to start off, however, by giving you a road map for the design of your study and showing you how your research question matters in the planning you'll be doing. Indeed, your initial choice of a research question will shape all other aspects of your proposed study. And as you revise that question, your plans for your data collection and analysis will undoubtedly change, too. That's normal for research design, and it is one of the reasons why the design process takes so long: you want to be confident that all the time and energy you put into a project will ultimately pay off.

Here is an outline of what the stages in research design might look like. We have also included the chapters in the textbook that will discuss each of these stages:

1. Develop an initial research question—what is called a **working research question**, given that it is a work in progress (**Chapter 4: Research Questions**)
2. Decide on an overall research strategy—whether your analytical approach will be inductive or deductive, what the target population is, and so on. (**Chapter 4: Research Questions**)
3. Conduct a literature review to identify a gap in the existing body of research (a **research problem**) and refine your research question (**Chapter 5: Research Design**)

4. Decide on a method of data collection and analysis (**Chapter 5: Research Design** and the chapters devoted to each method)
5. Propose hypotheses based on your literature review (**Chapter 5: Research Design**)
6. Decide on your sampling strategy (**Chapter 6: Sampling**)
7. Define key concepts and measures—what is called **conceptualization** and **operationalization**, respectively (**Chapter 7: Measuring the Social World**)
8. Identify and address any ethical concerns about your proposed study (**Chapter 8: Ethics**)

Please note that the stages of research design we've just described do not necessarily have to follow this order. For instance, you may have a good idea of the sample of research participants you want to recruit for your study well before you know the specific questions you'll ask of them. Also, you will continually move back and forth between steps—for example, between reading previous studies on your topic and further revising your research question, and between refocusing your research question and rethinking the ways you measure your variables. Nevertheless, at some point you will need to address all these stages of research design.

In this chapter, we'll start off by discussing ways of generating and refining a research question. We'll also cover different types of research questions, which roughly correspond to different types of data collection and analysis. Throughout the process of developing your research question, you want to be reviewing the existing studies on your topic, which is the best way to clarify and sharpen your question. This reading of previous research—what we call the academic **literature**—needs to happen throughout the research design process (and should continue even beyond it, to help make sense of your study's findings). We discuss how to write a literature review in **Chapter 5: Research Design**, but even after you move on to that stage, we advise you to keep revising your research question, using the advice in this current chapter to connect your question more clearly and creatively to the existing body of scientific knowledge on your topic.

Opening chapter image credit: *Olena Bohovyk, via Pexels. Adapted by Bizhan Khodabandeh.*

4.1. Choosing a General Research Topic



Don't be afraid to study what you are passionate about—whether it's a hot topic in the news, or something that you are interested in, but that has received little attention so far. If you are part of a group (from a membership group like a fraternity to an identity group like the LGBTQ+ community), you can take advantage of your insider perspective and contacts to help you research it. It's also fine to study communities you don't belong to. Either way, you will need to be sensitive to the various ethical issues that arise from being too detached or attached to the subjects of your research (which we discuss at length in later chapters). *Gotta Be Worth It*, *via Pexels*

Learning Objectives

1. Learn how to identify a topic to investigate.
2. Learn how to ask questions to narrow your research topic.
3. Learn how to create concept maps that demonstrate the relationships between concepts relevant to your topic.

How do you identify a good topic to study? Thankfully, sociology is such a broad field that we sociologists can study virtually anything we want, as long as it has some connection to society and social life. For that reason, we'd encourage you (with your instructor's approval, of course) to not limit yourself to what you think other people want you to study. Think about whatever excites and energizes you. Which social theories really speak to you? Which social issues are you passionate about? Which public policies—existing or proposed—do you see making a significant difference in people's lives? Which communities or groups do you want to help? Use these questions to identify a general topic that interests you enough to spend a substantial amount of time reading and writing about it.

How do you know your topic is relevant to sociology? One easy way is to think about the sociological theories you know (including the major paradigms and theoretical frameworks we discussed in **Chapter 3: The Role of Theory in Research**). Can you imagine any of these theories being relevant to the subject you're studying? If so, your study has the potential to be sociological—even if none of the people currently studying this phenomenon or issue are sociologists.

Here are some other suggestions for picking a research topic:

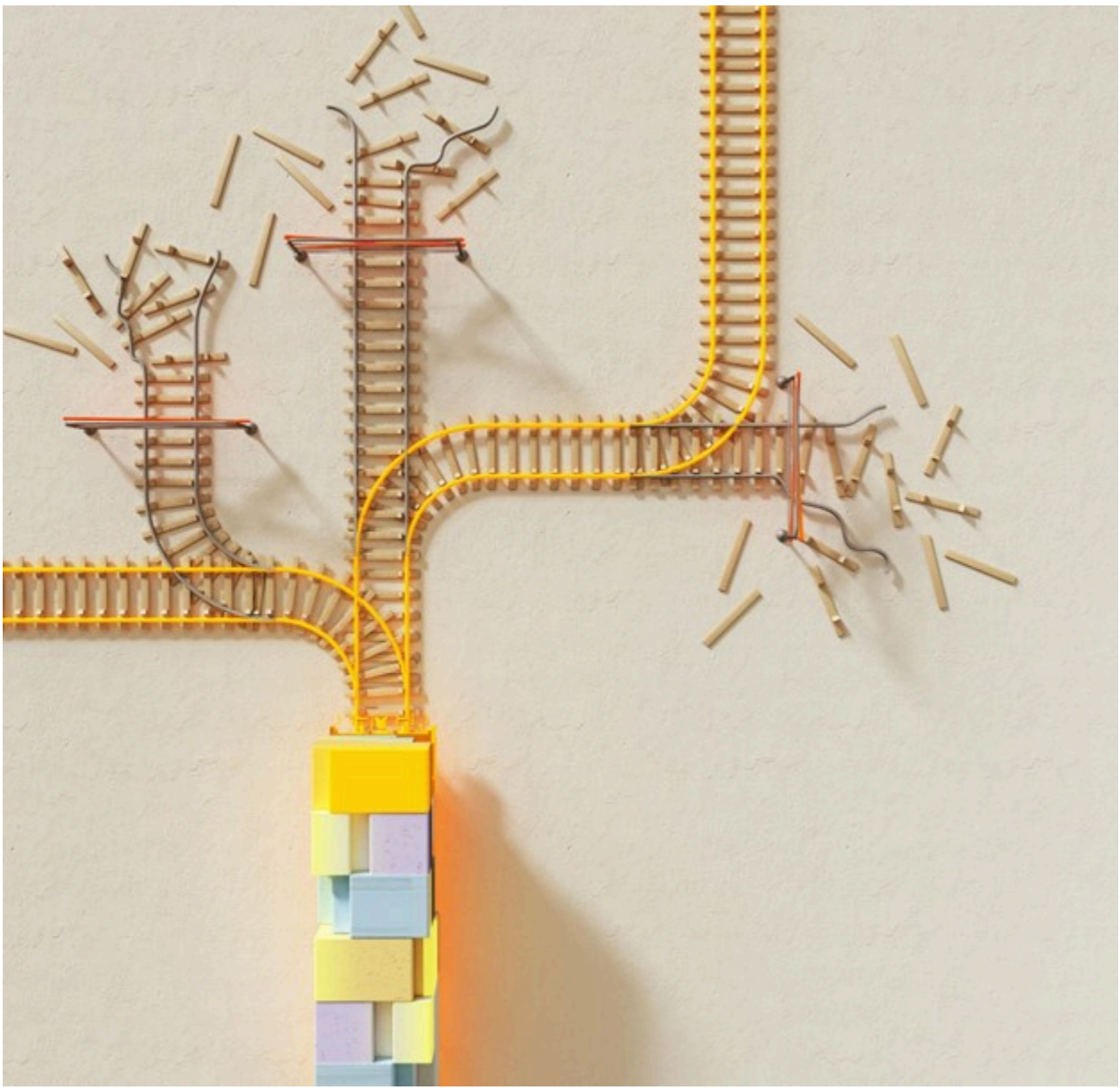
1. **Choose a topic based on the areas where you have noteworthy experience—professional, personal, or academic.** Think about papers you enjoyed researching and writing for other classes. Consider a specific group that you've interacted with or just care a lot about—for example, youth who identify as LGBTQ+ or patients at a community health clinic. Start with a social problem you've encountered before, such as gang violence or cyberbullying, or a policy or program you've experienced firsthand, such as healthcare regulations or zero-tolerance policies in schools.
2. **Think about what you're intellectually curious about.** Maybe you'd like to read up on specific sociological theories—start doing that, especially any academic writing on the subject. That should generate some ideas. If you can find out—or already know—whether more research is needed in a particular area, you'll make your task of finding a good topic all the easier, because any new scientific study you propose ultimately needs to address such a research problem. To that end, you may want to ask a professor, preferably one active in research, about possible topics that are understudied in their subfield of specialty. You can also read up on what the faculty members and graduate students in your department or elsewhere are currently working on. That can tell you what the hotter topics of study are.
   Note that finding research ideas by talking to people and looking at what they are currently studying isn't necessarily "stealing." Plenty of scholars study the same topic at the same time. Furthermore, you'll find a collaborative, rather than competitive, spirit in many scientific communities that study a particular subject, given that scientists inevitably build on the work of their peers. (For that reason, too, we encourage you to share preliminary drafts of your research with other researchers when that time comes around.) As long as you credit other people for any ideas you borrow from them and do not just mindlessly mimic their research approach, your work should constructively complement those of other scholars.
   Faculty and graduate student research interests vary widely, and it might surprise you to know what topics they've published on in the past. Many departmental websites will post or excerpt the curricula vitae (CVs) of their faculty members, and both students and faculty often have professional websites and social media accounts where they describe their research and link to their CVs. (Longer than a résumé and used largely in research and teaching fields, a **curriculum vitae** lists a scholar's credentials, research interests, teaching experience, and important

publications and presentations.) You can read some of their work and maybe schedule a time to talk with them, which could spark further ideas. Many professors and graduate students are willing to chat about research even if you aren't taking a class with them—it helps, of course, if you have some existing connection to that person, such as being at the same school.

3. **Identify and browse academic journals related to your research interests.** **Appendix A: Presenting, Writing, and Publishing** lists a variety of sociology-related professional associations and the journals they publish. Faculty and librarians can help you identify relevant journals and find articles relating to your specific areas of interest. You can also skim the titles and abstracts of presentations at the conferences put on by professional associations (refer to that same list for links) to learn about the most cutting-edge research on a particular topic. As we describe in Appendix A, conference talks are usually about works in progress, so the topics they cover tend to be especially fresh and innovative.

If you pursue any of these kinds of brainstorming, eventually you'll come across research papers to read. (In scientific lingo, a **paper** is a write-up of a study that a researcher has conducted.) A paper's literature review (sometimes called a "background" section) will provide insight into the research question the author was seeking to address with their study. As you're reading the literature review, look for what's missing. Is the existing body of research on this topic incomplete, imprecise, biased, or inconsistent? If so, a gap exists in the literature that you might explore in your own study—assuming the author of the paper you're currently reading didn't address that very gap. Even if they did, they will usually flag additional research problems near the end of their paper—sometimes in stand-alone sections that explicitly talk about their study's limitations or future directions for further research, and sometimes just in the discussion or conclusion sections at the end.

Note that your existing views on a topic will undoubtedly influence how you study it, and it is important to be aware and transparent about any preconceived notions and biases you have. How do you feel when people have other views on this subject? Do you believe your perspective is the only valid one? If you feel so strongly about the topic, certain findings from your research might upset you. In that case, you might want to choose a different topic because you don't want to run the risk of even unintentionally designing a project that will get only the answer you believe to be best. That doesn't fit with the spirit of science, which is about seeking the truth (even if actually getting at "truth" is exceedingly challenging and complicated, as we discussed in the last chapter).



Be honest about any preconceptions or biases that you might bring into your research on a topic. You must be open to your empirical investigations taking you on a winding path that might end at an answer that you personally would prefer not to see. *Google DeepMind, via Pexels*

Likewise, as a researcher, you should not put yourself in the position of feeling any pressure to downplay findings you don't like. For example, imagine that you wanted to find out whether a relationship exists between intelligence and political party affiliation. You happen to believe (as many of us do) that the members of your favored party are without a doubt more intelligent, thoughtful, and attractive than their counterparts across the aisle. This deeply held belief of yours is not a problem in itself. If, however, you feel intense rage when considering the possibility that your study will actually find the opposing party's members to *not* be so inferior, you're probably too emotionally invested in this particular topic to conduct unbiased research on it. As a sociologist, you need to be open to being persuaded by the

conclusions of rigorous research, even those findings that may be inconsistent with or unflattering to your personal or political perspective.

If you are engaged in research that requires interacting closely with the people you're studying, you may find yourself encountering people with very different perspectives than your own. They may have strong opinions about the issue you're studying, and their beliefs and worldviews may seem strange or even repellent to you. Nevertheless, you as a sociologist need to represent their viewpoints faithfully in your research. That may mean stomaching views and actions you vehemently oppose to portray individuals or communities fully and fairly in your work. If you can take up this open-minded viewpoint—which not many people can—then you can do truly groundbreaking research. For example, Kathleen Blee has done fascinating work on hate movement participants, as seen in her books *Inside Organized Racism* (2003) and *Women of the Klan* (2008). She clearly does not share the racist ideologies of the people she talked to, but what comes across in her work is the honesty of her observations—even when she wholeheartedly disagrees with what is being said or done. If you can't bring a similar degree of patience and impartiality to your own research topic, then you may want to study something else.

Your feelings about a particular topic do not even have to be especially intense to influence your ability to research it. Scientists are well aware of the problem of **confirmation bias**—that we as researchers are inclined to interpret the data we collect in ways that support, or "confirm," our existing views. This is a major issue in research, which may help explain why scientists are finding out today that many past studies cannot be replicated (as we discuss further in **Chapter 12: Experiments**). Even well-meaning scientists can fall victim to *selective perception*, the unconscious tendency to see what we want or expect to see. As a social scientist, you would be wise to start off any study questioning what you already know about a topic—and allowing yourself to be surprised by what you actually find in the course of your research.

That said, having strong feelings or existing knowledge about a topic shouldn't necessarily stop you from studying it. Sociologists often find themselves motivated to conduct research that addresses inequalities or injustices. In academia, thinking a policy is wrong or a group is being marginalized does not necessarily mean that your research will be biased. For better or worse, the bar of objectivity that academics need to meet in their research is lower than that for, say, court judges or journalists at reputable news organizations, where members of those professions are discouraged or even prohibited from expressing their political beliefs.

Sometimes the best topics to study are those you do feel strongly about—what better way to stay motivated than to learn about something you care about? Furthermore, having a personal tie to a particular topic may make the actual work of data collection and analysis easier. For instance, if you happen to be a member of an organization, profession, or network you're thinking of studying, you'll probably find it easier to gain access to insider information and get people in the group to talk to you. Similarly, if you have directly experienced and intimately know a particular social phenomenon or problem, you may have the knowledge and skills to contribute more to the scientific study of it than any person new to that issue would be able to (see the sidebar **Asking Questions in Need of Answers**). Of course, you will have to weigh those potential advantages against the potential disadvantage of being too close to the group or experience you're studying to examine it accurately and critically.

A final word of advice about choosing a research topic: be mindful of your mental health. A sociologist who has experienced a mental health crisis or traumatic event should approach researching related topics cautiously. There are plenty of worthwhile things to study, and there's no need to jeopardize your well-being for a research paper.

Asking Questions in Need of Answers: A Q&A with

Joyce Rothschild



Joyce Rothschild is professor emerita at Virginia Tech, having retired after serving for 27 years as a faculty member in the university's Department of Sociology and School of Public and International Affairs. Her academic writings have covered topics ranging from worker cooperatives (companies whose employees own equal shares of the business and democratically make decisions on its behalf) to whistleblowers (people in organizations who divulge information about illegal or immoral activities within the group). A central question that has guided her research is how organizations can accomplish their purposes without top-down management. In a groundbreaking article published in the *American Sociological Review* (1979), Rothschild described a "collectivist-democratic" alternative to bureaucracy: organizations like worker cooperatives that allow all the group's members to help decide important matters and freely voice their opinions and preferences. Her book with J. Allen Whitt, *The Cooperative Workplace: Potentials and Dilemmas of Organizational Democracy and Participation* (1986)—which won the C. Wright Mills Book Award from the Society for the Study of Social Problems—is a qualitative study of how collectivist-democratic organizations navigate the challenges of managing their operations without a head boss or hierarchies. For her study, Rothschild engaged in **participant observation** (a method we describe in **Chapter 9: Ethnography**) by working alongside members of the organizations she was studying and seeing firsthand how they operated.[1]

Why did you decide to embark on a sociological study of organizations?

As a graduate student, I remember thinking that it was hard to affect large social forces or bring about political changes writ large. Yet with a limited number of people, one *could* build grassroots organizations that addressed major challenges. And these groups didn't have to be hierarchical and bureaucratic. Indeed, when I was in graduate school in the early and mid-1970s, organizations were popping up all around me that called themselves "collectives." In these groups, people banded together to achieve a substantive goal that they held in common. Members related to each other as equals or peers. The whole idea was to offer everyone a voice and get them involved in the decision-making process on an equal footing—things that conventional organizations did not do.

At the time, I was very taken with Max Weber's theoretical arguments about all the problems inherent to bureaucracies. This hierarchical and rule-bound form of organization dominated everywhere, even though many people found it to be incredibly stifling and oppressive. In other parts of the world, socialist countries had set up state-owned firms, but I could see that their bureaucratic organizations were not offering

1. You can read more about Rothschild's research and her take on the future of collectivist-democratic organizations in the interview "Participatory Democratic Organizations Everywhere: A Harbinger of Social Change?" (2021).

experiences to their workers that were any different from what capitalist bureaucratic organizations were offering here. Why bother, I thought, unless we can offer a *better* form of organization that gives people more autonomy, more voice, and more friendship-based relationships?

I decided to study these new "collectives," which were already resisting hierarchies and putting into practice more collectivist-democratic forms of organizing. They had sprung up, of course, out of the larger countercultural social movement that was growing at the time, but I chose to study them as newly emergent organizations with some lessons to offer.

What methodological challenges did you encounter in your research?

In any participant observational study, the researcher has to strike a balance. You need to share enough about your research to gain access to the setting and to be seen as a person who can be trusted and potentially helpful to the groups you are studying. At the same time, you don't want to share too much detail about how you may be seeing the situation. Otherwise, the people you are studying may just feed you what they think you want to hear. The same balance is needed when one embarks on a set of interviews. The right description of your project will get people to talk to you, but sharing too much information can bias what you learn from your respondents. Achieving the right balance is something that a new researcher should talk with their faculty mentors about.

A second challenge is time. You have to put in the time—whether it is an observational study or an interview-based study—to get the insight and in-depth knowledge you seek.

Students often struggle with specifying research questions. What do you think makes a research question "good"?

I think good questions are grounded in important theoretical questions. For example, When I started studying collectives, my central question was derived directly from the key question that Weber had asked almost a hundred years earlier: how do people organize on behalf of their collective interests? He argued that societies ultimately moved to the most efficient and rational form of organization, the bureaucracy, which operated according to formal hierarchies and impersonal rules. I saw something very different happening in collectives, and so my study's research questions explored *how* these organizations went about accomplishing various things. After I got to know firsthand how alternatives to bureaucracy actually worked, I ended up arguing that the world needed another way of organizing—a *post*-bureaucratic model.

This brings me to the second big thing that a good research question needs. To be significant, a study needs an empirical reference—hopefully, something that is newly arising in the world, or something that previous researchers have not noticed much. I could not have done a successful study as a thought exercise. Yes, I had a specific theory I wanted to explore, but once I entered the field, my job was to learn from actual data and look out for surprises. It was my subjects who came up with new practices and a new rationale for their new organizational form—not me. I just observed and tried to faithfully describe what I saw.

Do you have any general advice for coming up with good research questions?

When I took methods courses as a graduate student, we were told that we should avoid being very involved with our subjects of study, as that might bias our research. However, more often than not, researchers do draw from their own experiences and attachments. This is what leads them to choose their area of study. It's what has catalyzed their commitment. It's how they know about an understudied phenomenon in the first place. I would say don't be afraid to draw upon your own experiences and concerns as you delve into the process of figuring out what you most want to study. Once you decide you don't want to just replicate and write about what is already well-known, then your only comparative advantage in research, as I see it, comes from your own involvements in life. This is what enables you to see something that others may not have noticed.

This is why my advice to the graduate students I have mentored over the years has been to scour your own life for experiences that have been meaningful to you. These experiences may have aroused great anxieties for you. Or they may have provided you with much-needed calm. Regardless, they raised questions—questions for which you presently have no answer. So, this is my answer to the problem of research bias: don't avoid involvement with your subject matter, as involvement is what you'll need to discover an important question that has not already been overworked. Instead, ask a question that your subjects need the answer to. That is what will lead you to an important research question. Remember, this is not just about your need to write a thesis. It is about other people's need to learn some answers that will help them navigate an important situation they're facing. Design your research project around that question, understanding that your answers will be unearthed in the process of your research.

In my case, I knew these new "collectives" were important insofar as they were seeking to do something that Weber had said was impossible in a modern society: building organizational realities that were peer-based, egalitarian, and democratic, not hierarchical or rule-bound. But how could they make their workplace processes and their decision-making processes work? The answer to this question was important to my subjects—to the many people whose organizations I was studying—and this is what motivated me and kept me going through the long process of research that ultimately gave rise to my book, *The Cooperative Workplace: Potentials and Dilemmas of Organizational Democracy and Participation* (1986).

Narrowing Your Research Topic



At some point, you will need to move from a research topic to a research question. Focus on a particular aspect of the broad topic you are studying, thinking about relevant concepts and variables and how they relate to one another. Your brainstorming will eventually bring you to a research question. *Suliman Sllehi, via Pexels*

Once you have a topic, think about the different aspects of it, and what specific angle might be interesting to pursue for your study. What do you really want to know about the subject? As a warm-up exercise, try dropping a possible topic idea into the blank spaces of the questions in the following list. You'll start thinking about the different dimensions of the topic you're studying and what exactly piques your intellectual curiosity about it.

Defining and differentiating a phenomenon:

1. What does ___ mean? What is the significance of ___? What is the value of ___?
2. What are the various points or features of ___? What are the facts about ___?
3. What are the component parts of ___? What are the types of ___? What kind of person is ___?
4. How is ___ like or unlike ___?
5. What case can be made for or against ___?

Understanding a process relating to the phenomenon:

1. How is ___ made or done? How should ___ be made or done?
2. What is the essential function of ___?
3. What are the causes of ___? How did ___ happen?
4. What are the consequences of ___? What is the relationship between ___ and the outcome of ___?

Understanding the context in which the phenomenon occurs:

1. How is ___ different in ___ environment, time period, or other context?
2. What is the present status of ___ (compared with ____)?

The third set of questions gets at another important aspect of a good research question: it starts with a specific population in mind. We want to know something about our phenomenon within a particular human population (e.g., residents of one country, all students, all members of a particular ethnic group, all human beings) or nonhuman population (e.g., colleges and universities in one country, all books published in a particular year, all multinational corporations). We'll have more to say about our study's **population of interest**, or **target population**, in **Chapter 6: Sampling**, but even at this early point, we would still advise you to have a specific population in mind when you devise your working research question. For instance, you might phrase your research question in this way: "This study examines how Policy X affects Outcome Y in Population Z." (Note that the research question does not have to be phrased as a literal question—in fact, it often sounds better if you don't do that!)

Developing a **concept map** around your topic may also help you narrow your research topic and generate a working research question. Using this technique, start with the broad topic, issue, or problem, and begin writing down keywords—all the words, phrases, and ideas related to that topic that come to mind. See which keywords can be grouped into larger concepts, and which concepts can be broken down into different types or aspects of that concept. Put all these keywords in boxes or circles and connect any related items with arrows. The arrows should point from what you think might be an independent variable toward what you think might be a dependent variable. Of course, many relationships are actually bidirectional: for instance, how much education you have affects your income, but your income can also affect how much education you attain. For the sake of brainstorming, however, you may want to focus on a particular direction—a particular cause-and-effect sequence. If education is your independent variable, for

example, then focus on the arrow that goes from that box to the dependent variable of income: in this case, you have set yourself down the path of exploring research questions like how higher education matters in determining people's later income. This approach of identifying concepts and connecting them causally will allow you to keep adding more boxes and arrows until you have a broad sense of how all the concepts that are relevant to your proposed study are linked by a chain of possible cause-effect relationships. You may then decide to frame your research question around one or more of those relationships, as we discuss in the next section.

In addition to assisting you in developing a research question, concept mapping will give you some keywords to use when searching the literature for relevant studies and publications on your topic (which we discuss in **Chapter 5: Research Design**). Concept mapping can also be helpful when you actually start writing your literature review, as it demonstrates the importance of each concept and its subconcepts, as well as the relationships between them (**Figure 4.1**).

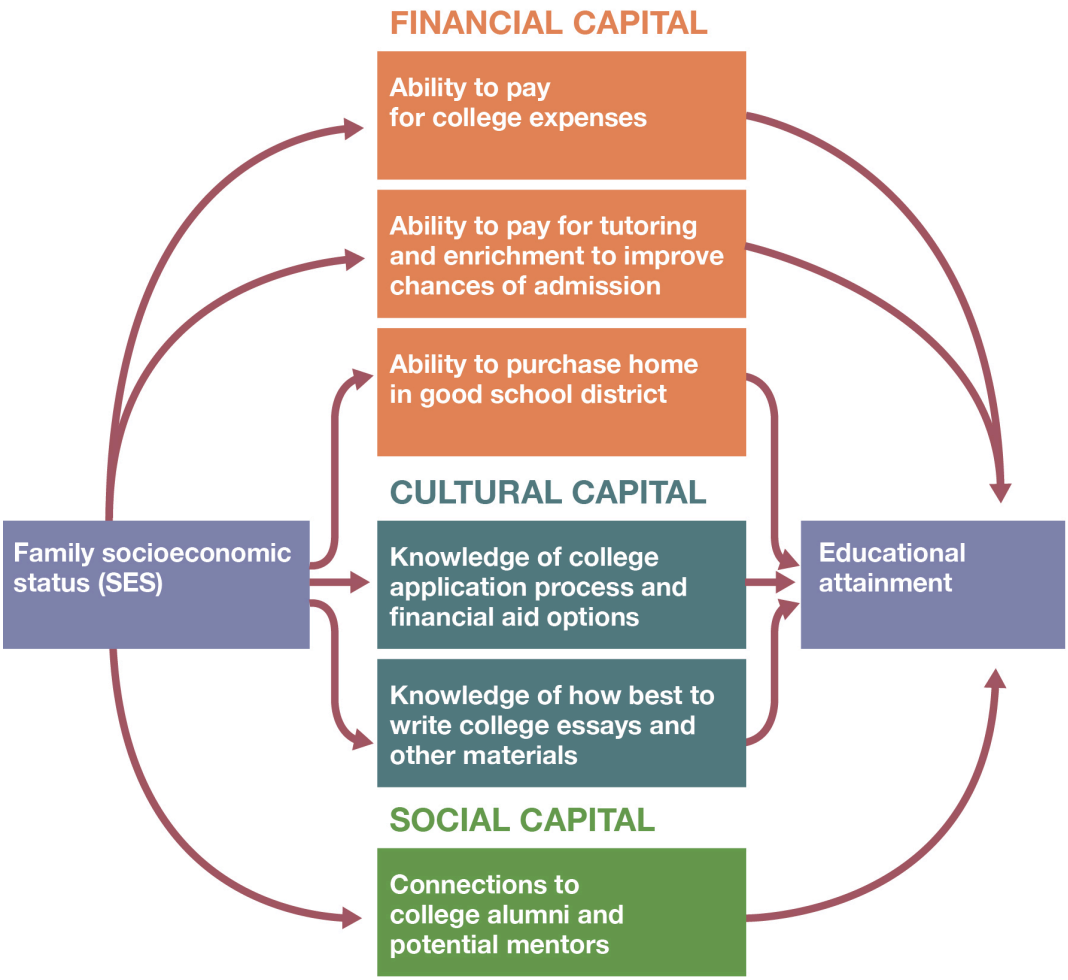


Figure 4.1. Example of a Concept Map. To brainstorm a research topic and research question, start with some general ideas about what you might be interested in and create a concept map that identifies any relevant concepts and the possible connections between them. Break down broad concepts into more specific ones, and see how related concepts could represent a larger phenomenon. For instance, let's say we were interested in all the ways that family advantages help students get a college degree. In this case, a

key relationship would be between family socioeconomic status (the parents' income, education, etc.) and educational attainment (whether their children go on to attend college, etc.). But we might also brainstorm the specific ways that family socioeconomic status (SES) leads to more education for the children—what we call *causal mechanisms*. Off the top of our heads, we might write down the mechanisms listed in this map: through these pathways, we hypothesize, family SES affects educational attainment. Then we might think how these mechanisms are related—for instance, that the first three have to do with financial capital (how much money the family has), the second two have to do with cultural capital (the useful cultural and professional knowledge that children receive from their households and social networks), and the last has to do with social capital (advantageous ties to other people).

So far, we've started asking questions about our topic and sketched out a concept map describing aspects of it. Now what? The next step is to actually draft a working research question. Take a minute right now and write down a question you want to answer with your research. Even if it doesn't seem perfect, everyone needs a place to start.

Next, try to narrow down your question. Continue to use the list of questions we provided earlier to help you in that task. Also single out the boxes you want to focus on in your concept map: after taking any overly broad concepts and breaking them down into different types or aspects, you should flag which of them you're most interested in. Eventually, you will settle on one or a few key relationships—that is, pairs of independent and dependent variables (and perhaps moderating or mediating variables that influence those relationships, as discussed in **Chapter 3: The Role of Theory in Research**).

For example, perhaps your initial idea is to study obesity. After an initial search of the relevant literature, you realize the topic of obesity is too broad to adequately cover in the time you have to do your literature review. You decide to narrow your focus to the causes of childhood obesity. Using the list of questions and your own concept mapping, you further narrow your search to the influence of family factors on overweight children. You are particularly interested in the ways that parents oversee their children's diets: whether and how they prepare or purchase meals, sit down for family dinners, introduce vegetables and other healthy options, and so on. You also decide that toddlers in the United

States are the target population you wish to study. Your working research question might then be "How do parental decisions about organizing meals affect toddler obesity in the United States?"

Key Takeaways

1. Many researchers choose topics by considering their personal experiences, knowledge, and interests, which may make it easier or more personally fulfilling to conduct a particular line of research.
2. You should be aware of and forthcoming about any strong feelings or preconceptions you have about your research topic, which may affect your ability to conduct unbiased research.
3. There are benefits and drawbacks associated with studying a topic about which you already have some prior knowledge or experience, so you should weigh those pros and cons when deciding on the focus of your research.
4. Ask questions about your research topic and sketch out a concept map that includes related keywords. This will help you focus your topic on a manageable set of variables that you find interesting to examine.

4.2. Types of Research Questions

Learning Objectives

1. Define empirical and normative questions and provide examples of each.
2. Understand the differences between exploratory, descriptive, and explanatory studies and research questions.

As you move from a research topic to a research question, some considerations should guide how you pose your question. First, social scientists are best equipped to answer **empirical questions**—questions about the facts of the world around us—as opposed to normative or ethical questions—questions about what we as a society should value. Empirical questions can be answered through research, but the answers to **normative questions** depend on people's moral opinions. (To say something is "normative" means that it relates to our norms or standards—what we *should* do.) While research projects can inform how we make decisions about ethical issues, they cannot directly answer normative questions, which are fundamentally a matter of debate within communities and societies about what sorts of principles they want to uphold.

For example, a student in one of our methods classes wanted to research student athletes. Their original research question was: "Should college athletes be paid?" Outside of a research context, this is a great question—the matter of paying or not paying athletes affects the lives of millions of students, and it speaks to critical issues about what we as a society think a college education should entail, and what is a fair reward for the work people do. Unfortunately, this specific question is a normative one that we need to debate, not an empirical question that we can resolve with research. The tip-off is that it begins with the word "should," a normative phrasing that you generally want to avoid in research questions. The answer to such a question would be a series of moral arguments, based on the particular values the author and their audience hold in common.

It's true that research can help us to make moral arguments. For example, if we learn how much money universities make from college sports, or how all the work that athletes put into training and playing shapes their experience of college, that empirical knowledge could help us decide whether we believe student athletes are being exploited by their universities, and whether we believe they have a moral right to be paid for their labor. But then *those* questions would be our research questions, rather than the normative question of whether athletes "should" be paid.

Let's consider another ethical question that research can inform but not answer: is *SpongeBob SquarePants* immoral? In 2012, a Ukrainian government commission began reviewing that cartoon show in response to complaints by a right-wing religious group that its depiction of depraved behaviors—such as SpongeBob's regular practice of holding his male sidekick Patrick's hand—amounted to the "promotion of homosexuality" (Marson 2012). (Before the government body was disbanded in 2015, the National Expert Commission of Ukraine on the Protection of Public Morality evaluated media to ensure that television shows and other content adhered to the country's morality laws regarding pornography and other controversial issues.) The agency called a special session to discuss *SpongeBob* and other suspect kids' shows, though ultimately the eponymous sponge and his starfish companion stayed on Ukrainian TV. A decade earlier, SpongeBob had also drawn the ire of U.S. conservative groups for appearing with other popular cartoon characters in a music video intended to teach children about multiculturalism—which the advocacy group Focus on the Family said was

"pro-homosexual" and served as an "insidious means" of "manipulating and potentially brainwashing kids" (Kirkpatrick 2005).



Research questions must be empirical questions, which means you can answer them by collecting and analyzing data. Normative questions like, "Is SpongeBob SquarePants immoral?" are rooted in people's values and can only be informed by research, not answered by it. That said, a study could answer the empirical question of how many Americans believe SpongeBob is immoral. *Nickelodeon, via IMDb*

Can research answer the question of whether *SpongeBob SquarePants* is immoral? No, because questions of morality are ethical, not empirical. Your family members and pundits on TV can rant about sponge creatures all they want, and they can make better or worse moral arguments for their positions, but this is not a question a social scientist should build a study around. That said, we sociologists could certainly choose to study the public opinions and cultural meanings that surround a popular show like *SpongeBob SquarePants*. We could conduct experiments measuring the detrimental effects that watching the show has on children's behavior. We could even use surveys to find out precisely how many people in the United States find SpongeBob and/or Patrick repugnant. But sadly, we could not settle the question of whether SpongeBob is indeed morally reprehensible, given that it is not an empirical question.

As you start designing your study, your choice of a particular empirical question will also be influenced by your study's general purpose. There are three approaches that a research study will typically take: *exploration*, *description*, or *explanation*. These categories are not mutually exclusive, and a study may fall into multiple categories.

Exploratory research is often conducted in new areas of inquiry, where the goals of the research are: (1) to scope out the magnitude or extent of a particular phenomenon; (2) to generate some initial ideas or hunches about that phenomenon; or (3) to test the feasibility of undertaking a more extensive study. For instance, if the citizens of a country are generally dissatisfied with their government's policies during an economic recession, sociologists could create and implement new surveys to measure the extent of that dissatisfaction and probe for possible causes of it, such as anxieties about unemployment, inflation, or higher taxes. This research may not lead to a very accurate understanding of the target problem, but it may be worthwhile nonetheless to get a preliminary sense of its nature and extent, serving as a stepping stone to more in-depth research.

Descriptive research is directed at making careful observations and detailed documentation of a phenomenon of interest. Because these observations follow the scientific method, they hopefully are more accurate than casual observations by untrained people. Much exploratory research overlaps with descriptive research: we often want to describe the magnitude of an emerging problem as a starting point in understanding it. Yet descriptive research is also helpful to conduct on an ongoing basis, and it can involve well-studied topics.

A common type of descriptive research is the work of government agencies to tabulate statistics about the population. In the United States, for example, the Bureau of Labor Statistics uses survey questions to estimate employment by sector every month. The U.S. Census Bureau regularly conducts demographic surveys that allow policymakers and social scientists to track the growth of a wide range of racial and ethnic groups over many years. In general, government agencies, corporations, nonprofits, and other organizations are in great need of such descriptive research so that they understand the circumstances that their citizens, clients, and members are experiencing. They can use these assessments to create new programs or policies to meet people's needs or preferences. For that reason, if you decide to use your sociological research skills in a nonacademic setting (as described in **Chapter 2: Using Sociology in Everyday Life**), you will likely be doing a lot of descriptive research.

That said, descriptive research is also a large component of many studies that academic sociologists do. For example, a sociologist's ethnographic study of gang activities among adolescent youth in urban areas might entail detailed observations of the children's activities. A study of religious practices in immigrant communities might chronicle the evolution of those practices over time. In conducting this descriptive research, sociologists are gathering essential information about what is actually going on in the social spaces they observe.

Explanatory research seeks explanations of observed behaviors, problems, or other phenomena. You might think of the difference between descriptive and explanatory research in this way: the former seeks answers to basic "what," "where," "who," and "when" types of questions, whereas the latter examines questions that are more complex—such as whether or not one concept affects another, and "why" and "how" those concepts are related. Put another way, explanatory research attempts to "connect the dots" in research, by identifying certain important factors and showing how they lead to specific outcomes. Let's consider the hypothetical studies we discussed at the end of the last paragraph in this light. For an explanatory study about urban gangs, sociologists might seek to understand the reasons that adolescent youth in urban areas get involved with gangs. For an explanatory study of immigrant religious practices, researchers might examine why these practices evolve in the ways they do within particular local or national contexts.



Most scholarly studies mix elements of descriptive and explanatory research. For instance, a qualitative study of a street gang might involve a detailed chronicling of how members of the gang are recruited and what they do as part of the group. In addition to this descriptive data, however, academic sociologists would likely be interested in explaining some aspect of the organization, too, with the hope that the explanation they arrive at will apply not just to this specific group but also to others of a similar type. For instance, they might seek to explain why young people join the gang being observed, or why their ties to the gang supersede their connections to their other family and friend networks. *Khoa Võ, via Pexels*

Most studies you read in the academic literature will be explanatory. Why is that? Explanatory research tries to identify *causal* relationships that are generalizable across space and time. That means the findings of such research should matter to many people: because we're learning something fundamental about the relationships between the concepts we're interested in, our conclusions aren't limited to a one-off situation or scenario. It also means our findings are actionable: because we know what causes what, we can act individually or collectively to promote, discourage, or alter the phenomenon we're studying. In other words, explanatory research gives us a better sense of how and why society operates the way it does, rather than just describing what particular aspects of society look like.

Arriving at compelling explanations for social phenomena requires especially strong theoretical and empirical skills. You need to have a sophisticated understanding of how a social process operates and rule out any alternative stories, and you need to collect empirically sound data and rigorously analyze it. For these reasons, sociologists often see explanatory research as a "higher" form of research, one that is exceedingly challenging to do well. At the same time, they will frequently engage in some amount of exploratory and descriptive research for any given study, particularly during its initial phases. Indeed, these other approaches can be especially important in helping us understand a relatively new or hard-to-study phenomenon: without good descriptive research to draw from, any theorizing we do will be built on shaky empirical foundations.

Deciding on the primary purpose of your research will shape the study you ultimately propose and conduct. If you are doing academic work, your instructor or advisor may push you to be less "descriptive" in your approach and to focus more on seeking explanations for what you observe. If you are studying a topic that so far has generated only a small amount of literature, however, you may very well want to conduct exploratory research to generate plausible theories, or descriptive research to understand the scale or characteristics of a particular phenomenon.

The overall purpose of your research will also inform the research questions that you pose. Probably the easiest questions to think of are descriptive research questions. For example, "What is the average student debt load of college graduates?" is a descriptive question—and an important one. In this case, you aren't trying to identify a causal relationship. You're simply trying to describe how much debt students carry. When you seek to answer a descriptive research question like this one, you might find yourself generating **descriptive statistics**—counting the number of instances of a phenomenon, or determining an average, median, or percentage. You can also pursue descriptive research questions using qualitative methods. For instance, you might conduct in-depth interviews or focus groups to gauge the public's view of student debt, describing the range of opinions on that subject.

In the next section, we'll focus on explanatory research questions. We will detail one strategy for developing questions based on whether your study is using quantitative methods or qualitative methods. We'll also connect those two types of research questions to two kinds of empirical analysis—deduction and induction.

Key Takeaways

1. Empirical questions can be answered by gathering and analyzing data. Normative questions have to do with people's moral values and opinions and can only be informed, but not answered, through empirical research.
2. Exploratory research focuses on tentatively understanding new and emerging phenomena by gathering details and formulating plausible theories. Descriptive research involves a careful measurement of what a phenomenon looks like. Explanatory research tries to understand whether and to what extent two

concepts are causally related, and how and why they are related.

3. Descriptive questions are helpful for assessing current conditions for policy implementation and other purposes, but they do not investigate causal relationships between variables, which social scientists are often interested in.

4.3. Posing Explanatory Research Questions

Learning Objectives

1. Describe some common ways that sociologists pose explanatory research questions for qualitative and quantitative projects.
2. Distinguish inductive and deductive approaches to empirical research, and provide examples of each.
3. Understand how induction and deduction complement each other.

As we mentioned, explanatory research focuses on *causal relationships*—whether and how a change in one variable causes a change in another variable. As a result, explanatory research questions should contain an independent variable and dependent variable. The point of explanatory research is to determine the relationship between these variables (which, technically, are really concepts at this stage before we have specified how we will measure them—at which point they become **variables**). Indeed, as we noted, when you are mulling over the concept map you created earlier in the chapter, you can devise a simple research question by singling out two boxes—two variables—listed on your map. An explanatory question could involve the relationship between any two of your boxes, which would be captured by the direction of the arrow connecting them—that is, the direction of the causal relationship that you believe exists between those variables. (For more context, refer to our discussion of causality in **Chapter 3: The Role of Theory in Research**.)

Broadly speaking, the way that we pose an explanatory research question will be different when we are conducting a qualitative study and a quantitative study. One common format for an explanatory *quantitative* research question is this:

> What is the effect of the independent variable on the dependent variable within the target population?

A common format for an explanatory *qualitative* research question is this:

> How (or why) does the independent variable affect the dependent variable within the target population?

We should point out that plenty of quantitative and qualitative studies phrase their research questions in different ways than what we've just suggested. We advise novice researchers, however, to use these templates for a couple of reasons. First, as we described in **Chapter 3: The Role of Theory in Research**, quantitative research tends to be better at assessing *whether* a causal relationship exists between two variables within the target population—and if the relationship does exist, it can quantify the strength of that relationship (i.e., *the extent to which* one variable affects another). In contrast, qualitative research excels at identifying and describing causal mechanisms—the specific pathways or processes by which one variable affects another—and, in doing so, it gets at the *how* and *why* questions of that causal relationship. Posing explanatory research questions in these two ways therefore plays to the strength of each methodological approach: the ability of qualitative methods to generate complex data and nuanced explanations, and the ability of quantitative methods to isolate key variables, assess whether a cause-effect relationship between them exists, and estimate the size of that effect (if any) within the target population.

Furthermore, these two templates for research questions correspond—again, generally speaking—to the two major analytical approaches that sociologists take when conducting research: *deduction* (which is often the approach of quantitative studies) and *induction* (which is associated more with qualitative studies). When researchers take an **inductive approach**, they start with a set of observations and then move from those particular experiences to a more

general set of propositions about how the world operates. In other words, they move from data to theory, or from the specific to the general. Researchers taking a **deductive approach** take the steps just described and reverse their order. They start with a social theory that they find noteworthy and then test its implications with data. In this way, they move from a more general level to a more specific one.

In deductive research, sociologists are usually building explicitly upon what others have done—that is, reading existing theories of whatever phenomenon they are studying, and then testing hypotheses that come to mind when they apply those theories to specific social contexts. In inductive research, by contrast, sociologists are frequently generating new and original theories based on whatever patterns or regularities they've observed in social life, regardless of what past work has found (although, as we discuss later, most sociologists usually start their inductive investigations with some existing theories to expand on or contest). Put simply, inductive research is focused on *theory building*, and deductive research is focused on *theory testing*. Both approaches are critical for the advancement of science. Elegant theories are not valuable if they do not match with reality. Likewise, mountains of data are useless until they can contribute to the construction of meaningful theories.

Induction and deduction are often seen as two sides of the coin of scientific research, as you can see visually in **Figure 4.2**. Combined, they represent what is sometimes called the "wheel of science": a cycle of induction and deduction in which one approach picks up where the other left off. As this diagram implies, we scientists constantly shift between theorizing and making empirical observations—between induction and deduction—as we learn more about a topic. For instance, we engage in induction whenever we observe something and ask, "Why is this happening?" or "What broader trend is this a case of?" In answering these questions, we advance one or more tentative explanations. We then use deduction to test these hypotheses—in the process, perhaps narrowing down our list to the most plausible explanation, based on logic and our understanding of the phenomenon under study.

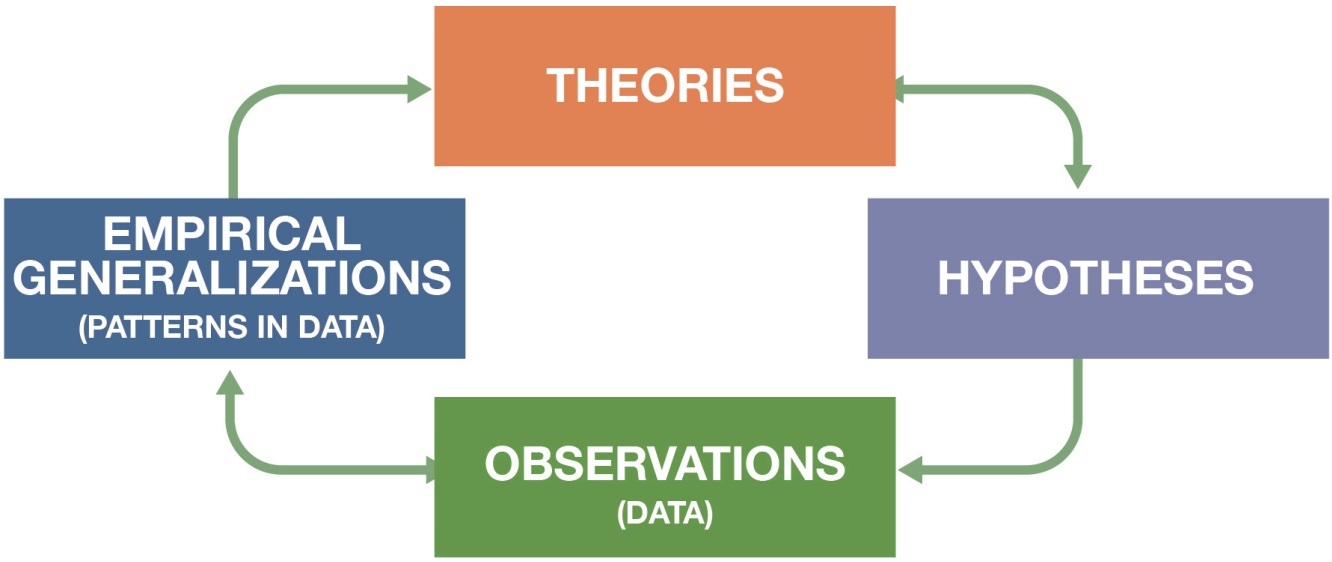


Figure 4.2. The Wheel of Science. As the wheel of science illustrates, inductive and deductive reasoning go hand in hand in the development and refinement of theories.

Let's go a bit deeper into each of these analytical approaches. We'll start with deductive approaches to test theories, and then go over inductive approaches to generate theories.

Deductive Approaches: Theory Testing

Deduction (**Figure 4.3**) is the version of scientific analysis that you were most likely taught in school or on educational television shows. In a deductive study, you start with a theory you want to test. You then apply that theory to a particular context (a particular population), developing one or more hypotheses—expected answers to your research question. Finally, you conduct data collection and analysis and see if your empirical results support your hypothesis. Again, the goal here is to *test* theories—use systematically gathered data to conclude with confidence that your hypothesis holds up or not. Ideally, you also want to go beyond just confirming or disconfirming the theory at hand, but also refining, improving, and extending it. Among other things, the results of your deductive investigation might allow you to revise your existing theory to say in what social contexts it is valid, and in what contexts it is not.

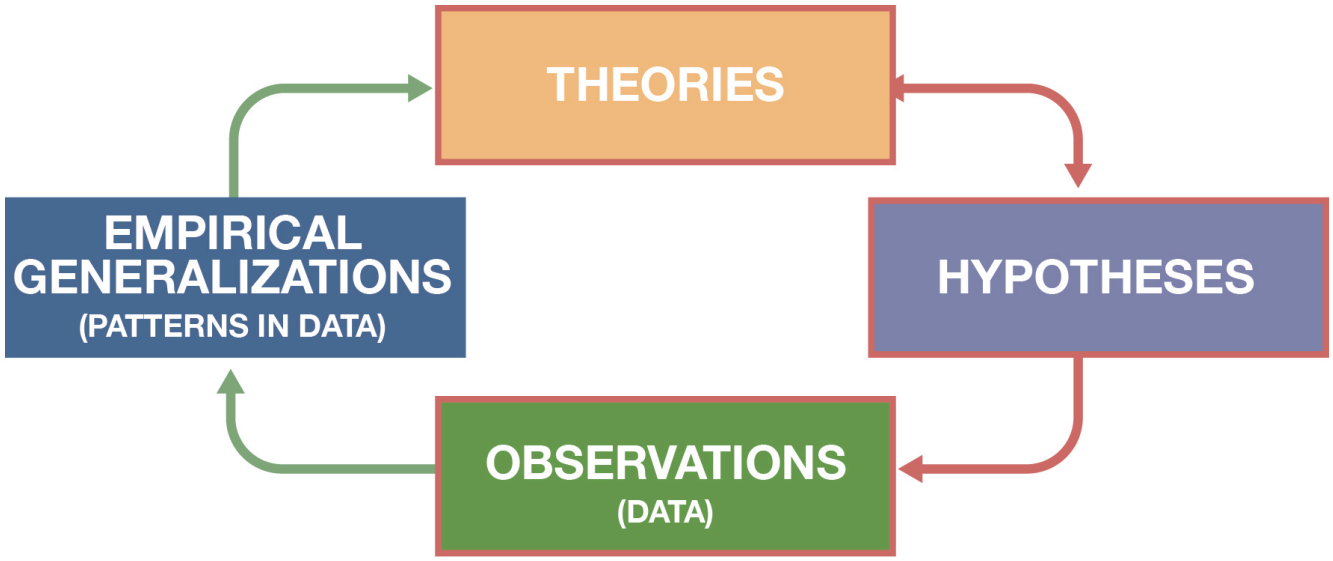


Figure 4.3. Deduction: From Abstract to Concrete. In the deductive stages of the wheel of science, researchers take an existing theory, apply it to a particular context to generate testable hypotheses, and then gather and analyze data to see if the hypothesized relationships hold up in the way that the theory suggests.

Let's consider some examples of deductive research. In their research on hate crimes, Ryan King, Steven Messner, and Robert Baller (2009) tackled a deductive research question: what effect does the locality where a hate crime occurs have on the response of law enforcement to that crime? Based on existing theories about the historical legacies of racial violence in different regions of the country, they hypothesized that law enforcement would respond less vigorously in areas where racial violence had prevailed during the Jim Crow era of segregation—areas that were largely, but not exclusively, in the South. They tested the hypothesis by analyzing county records of lynchings before 1930 and modern-day data from the same localities about how police responded to crimes falling under federal hate crime law—specifically, whether law enforcement complied with federal law, whether they reported hate crimes targeting African Americans, and whether they prosecuted hate crime cases. Overall, the authors found support for a modified version of their hypothesis: in areas where lynching prevailed before 1930 *and* where substantial populations of African Americans lived, police responded less forcefully to hate crimes.

In another well-done deductive analysis, Melissa Milkie and Catharine Warner (2011) studied the effects of different classroom environments on first graders' mental health (again, note the particular form that this *quantitative* research question takes: the *effect* of the independent variable *on* the dependent variable). Based on prior research, Milkie and Warner hypothesized that poor classroom environments—where students lacked basic supplies or even heat during the winter—would be associated with emotional and behavioral problems among those children. The researchers found support for their hypothesis, underscoring the fact that policymakers needed to pay more attention not just to children's grades but also to how their day-to-day experiences in school could affect their mental health.



Many decades after lynchings occurred in the U.S. South and other parts of the country, researchers were able to trace the legacies of that organized racial violence in terms of weaker enforcement of federal hate crime laws in the very same localities (King et al. 2009). *Victoria Pickering, via Flickr*

As social scientists ourselves, we are obviously biased, but we would venture to say that theory testing in the social sciences is generally harder than it is in the hard sciences. (For empirical support of this claim, we refer you to the tweet from physicist Neil deGrasse Tyson in **Chapter 1: Introduction**.) As we've previously discussed, the theoretical concepts that we social scientists use are often hard to define. The tools we have to measure those concepts—questionnaires, observations, and the like—have much less precision than, say, a spectrometer or oscilloscope, given how messy and multivariate human thinking and behavior is. We have to account for many more factors in the complex social worlds we study, all of which may also influence the phenomenon of interest. As a result, it can be very difficult to refute social theories that do not work, because proponents of a particular theory can always point to exceptions or dispute the ways that a phenomenon was measured or evaluated. For instance, for decades, countries persisted in organizing their economies according to Marxist theories of communism, in spite of mounting evidence that such approaches failed to promote economic growth and social welfare in a sustainable or even fairly equitable way. From the opposite ideological vantage point, the global financial crisis that spread from the United States to other parts of the world in 2008 failed to convince many policymakers to maintain tougher stands against reckless speculation and fraudulent profit-seeking pursued by Wall Street and the larger global financial industry—which have continued to engage in rampant risk-taking in recent years (Fligstein 2021). Sadly, even a preponderance of social scientific evidence often does not settle heavily politicized debates about important policies: every side can raise objections about the possible limitations of any research that doesn't favor their viewpoint, given the many potential shortcomings that are inevitably attached to any social scientific study, even well-executed ones. While few, if any, physicists dispute well-tested theories like gravity in any substantial way, plenty of social scientists would argue about the merits of social theories that have presumably been tested with decades of deductive research.

All this said, theory testing remains an essential step in social scientific inquiry. It gives us more confidence that we have arrived at the right story about how the social world works. Sociologists often favor a deductive approach because they feel it puts them on firmer empirical ground: their analysis of the data found (or didn't find) a relationship, the data came from a representative sample, and, therefore, the relationship exists (or doesn't exist) in the target population. End of

story. However, it is obvious that we need to start with a theory to be able to test it. That is where inductive approaches come in.

Inductive Approaches: Theory Generation

The other major kind of analysis used in scientific research is *induction* (Figure 4.4). Whereas deduction starts with the abstract—a broad theory—and moves to the concrete—empirical observations—induction does the opposite. In this approach, you start off by collecting data relevant to your topic of interest. Once you've collected a substantial amount of data, you then step back to get a bird's-eye view. You analyze the data to see what patterns emerge—for instance, how important variables appear to be related to one another within your sample, and how exactly these factors might be influencing one another. Based on these regularities (what are sometimes called **empirical generalizations**), you begin to piece together a larger theory that connects concepts in a way that's consistent with what you've actually observed. The theory you generate should help explain the patterns you saw in the data.

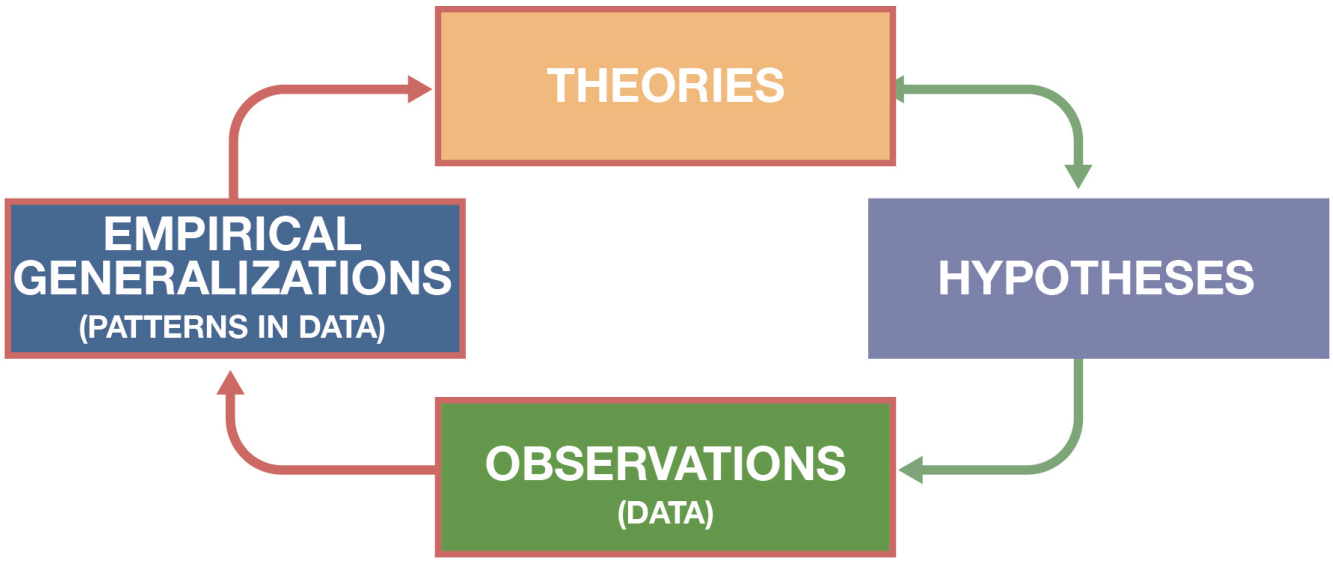


Figure 4.4. Induction: From Concrete to Abstract. In the inductive stages of the wheel of science, researchers make observations and identify noteworthy patterns and regularities in the data they gather. Based on these empirical generalizations, they generate concepts and identify relationships between them, thereby building and refining theories.

As we've noted, induction is often the approach taken by qualitative researchers. Pioneers of ethnographic methods used the term **grounded theory**—discussed further in Chapter 9: Ethnography—to describe their process of generating entirely new theories from a foundation of empirical data. The idea was that you would start with a clean slate—that is, having no theory in mind whatsoever, but letting the facts (data) lead you where they may. Nowadays, most qualitative researchers will consciously build upon or challenge existing theories from the outset (as we will describe later), but induction nonetheless remains central to their approach of data collection and analysis. For instance, Tomás Jiménez (2008) sought to understand how the process of assimilation changed for immigrant groups when the flow of newcomers into the country persisted (a context called "immigrant replenishment"), rather than petering away as it had for earlier generations of European immigrants. He conducted 123 in-depth interviews with second- and third-generation Mexican American immigrants in two cities, asking them about the sorts of incidents that made them feel more or less connected with their Mexican immigrant identities. Inductively, Jiménez identified two broad patterns

in his interviews. First, these later-generation immigrants felt their ethnic identities to be more salient when they had personal encounters, or heard local reports, of hostility against immigrants—incidents that were common in a context of immigrant replenishment. Second, many interviewees felt challenged as not "authentically" Mexican when they interacted with first-generation immigrants who spoke perfect Spanish and intimately knew the cultural practices of their home country—incidents also more likely to happen in a context of replenishment. These "intergroup" and "intragroup" tensions appeared to sharpen the boundaries that separated later-generation immigrants from other groups, perhaps explaining why assimilation had become more difficult in recent years.

We mentioned that deduction is what often comes to mind when people think about what science "is." Generally speaking, we can obtain more definitive answers to our research questions if we pursue a deductive approach (and use the quantitative methods that typically favor this approach). The questions tend to be more focused, and the samples more representative, which means that deductive studies often fit well with the scientific community's desire to have its work seen as rigorous and unimpeachable. These studies, however, are usually asking different sorts of questions, and as we'll discuss more in **Chapter 7: Measuring the Social World**, qualitative studies should be evaluated by different standards than quantitative studies are.



In his interview-based study of Mexican American assimilation, Tomás Jiménez (2008) used an inductive approach to understand why second- and third-generation immigrants in cities experiencing a substantial influx of new migrants felt their ethnic and racial identities to be more salient. While this strategy gave him a rich sense of the various ways that this context of "immigrant replenishment" influenced the identities of later-generation immigrants, he also was able to challenge existing theories—based on the experience of European immigrants in the nineteenth and twentieth centuries—that suggested immigrants inevitably became assimilated, with their ethnic identities eventually dwindling to the point of becoming merely "symbolic." For example, the intense anti-immigrant sentiment that the Irish encountered (illustrated in this racist political cartoon from 1871, "The Usual Irish Way of Doing Things," by Thomas Nast) also made their ethnic identities highly salient, but after the flows of new Irish immigrants dissipated, the ethnic boundaries separating the Irish from other white Americans also disappeared—which so far has not been the case for Mexican Americans, Jiménez argues, in part because of continued migrant flows. *Thomas Nast, via Wikimedia Commons*

Furthermore, a key strength of inductive research that we want to emphasize is that it doesn't lock us into one story about what is going on with a particular social phenomenon. As the fictional sleuth Sherlock Holmes said, "It is a capital mistake to theorize before one has data. Insensibly one begins to twist facts to suit theories, instead of theories to suit facts" (Doyle 2009). Because researchers doing inductive research begin by diving into the data, they open themselves up to possibilities they may not have considered at the outset. They can critique taken-for-granted assumptions and unearth previously ignored issues. They can absorb the stories they hear and observe and use that inspiration to theorize thoughtfully about any relationships among relevant concepts. They can easily find themselves surprised by what they encounter—and, in science, that's a good thing. Without this sort of research (which often fits into the *exploratory* category we described earlier), we're left with preconceived notions that still may not reflect the current state of affairs. It reminds us of the joke about the drunk man who dropped his car keys on the street at night and scoured the sidewalk looking fruitlessly for them. When a passerby asked why the man was only looking for them under a nearby streetlamp, and not the dark sections of the sidewalk, he answered, "Because the light is better here." Deduction helps us better understand what we already know; induction lets us explore in the dark.

How Inductive and Deductive Approaches Complement Each Other



Most students learn that the scientific method is a deductive approach that starts with theories, proposes hypotheses, and tests those hypotheses with data. Scientific progress also relies, however, on an inductive approach that creatively generates the theories that deduction can then test. *whiteinshiningarmor, via Reddit*

As we have emphasized, induction and deduction are each useful in their own way. Each has limitations. And each builds upon the other, helping bring us to better theories. Inductive research is especially valuable when few theories exist to explain a particular phenomenon, whereas deductive research is more productive when many competing theories are at play, and researchers want to know which works best, and under what circumstances. Induction identifies new patterns, such as relationships between variables, but we typically need to do more to say with confidence that those relationships are causal or hold up in the target population (specifically, we need to gather a representative sample and rule out any alternative causal stories). Likewise, deduction can help us verify that a theory actually applies to the population of interest, but we need to remember that other factors we haven't fully considered—other theories yet to be generated—may also be at work, and may actually be more decisive.

Indeed, rather than viewing these two processes in a circular relationship, perhaps they can be better viewed as a helix, with each dialectical iteration between theory and data contributing to better theories—that is, better explanations of the phenomenon of interest. Ultimately, we scientists must be able to move back and forth between inductive and deductive reasoning if we are to extend, modify, or overturn a given model or theory—which is how science advances. To that end, sociologists at times will design their studies to include multiple components, one inductive and the other deductive. In other cases, they might begin a study with the plan to only conduct either inductive or deductive research, but then they discover along the way that the other approach is needed to illuminate their findings.

An example of the latter situation is a classic study by Lawrence Sherman and Richard Berk (1984) of the effects of punishment on deterring domestic violence. The researchers decided to examine domestic violence as a form of socially deviant behavior. Based on existing sociological understandings of deviance, they sought to test two theories that seemed to explain whether an accused spouse batterer would respond to an arrest by engaging in more or less abuse. Deterrence theory predicts that arresting accused batterers will *reduce* future incidents of violence because they seek to avoid future punishment. Conversely, labeling theory predicts that arresting accused batterers will *increase* future incidents because the act of arresting them compels them to internalize that deviant identity. **Table 4.1** summarizes the two competing theories and the predictions that the researchers set out to test. To see which theory provided a better explanation, Sherman and Berk conducted an experiment with the help of local police in one city. They found that arrests were in fact followed by fewer incidents of reported violence—in line with the predictions of deterrence theory.

Table 4.1. Predicting the Effects of Arrest on Future Spouse Battery: Testing Theories through Deduction (Sherman and Berk 1984)

Theory	Prediction
Deterrence theory	More arrests → more incidents of domestic violence
Labeling theory	More arrests → fewer incidents of domestic violence

After conducting this study, Sherman and Berk and other researchers went on to conduct similar experiments in six other cities (Berk et al. 1992; Pate and Hamilton 1992; Sherman et al. 1992). Results from these follow-up studies were mixed. In some cases, arrest deterred future incidents of violence. In other cases, it did not. This left the researchers with new data that they needed to explain. The researchers therefore took an inductive approach to make sense of their latest empirical observations. The new studies revealed that arrest seemed to have a deterrent effect for those who were married and employed, but that it led to increased offenses for those who were unmarried and unemployed. The researchers thus turned to *control theory*, which predicts that the social ties provided by marriage and employment give certain accused batterers a greater stake in conformity. **Table 4.2** summarizes the propositions that control theory puts forward about the relationship between arrests and future incidents of violence, which the theory predicts will shift depending on whether accused batterers are married and employed. (To use the lingo we described in **Chapter 3: The Role of Theory in Research**, marriage and employment here are **moderating variables**, altering the relationship between arrests and future incidents of domestic violence.) Inserting control theory into their analysis provided the researchers with a superior explanation that better fit the complex nature of the relationships being studied.

Table 4.2. Predicting the Effects of Arrest on Future Spouse Battery: A New Theory Generated from Induction

Theory	Prediction
~~Deterrence theory~~	~~More arrests → more incidents of domestic violence~~
~~Labeling theory~~	~~More arrests → fewer incidents of domestic violence~~
Control theory	More arrests → fewer incidents of domestic violence for the married and employed More arrests → more incidents of domestic violence for the unmarried and unemployed

As Sherman and Berk's original paper and their follow-up studies show us, we as social scientists need to be willing to adapt our analytical approach to the question that is currently before us. In their case, they started with a deductive approach to research but were later confronted with new data that they had to make sense of. In response, they moved to an inductive approach.

We want to wrap up this discussion of analytical approaches by talking about how qualitative and quantitative research approaches figure into this back-and-forth movement between induction and deduction. As we've noted, induction is focused on generating theory—developing a model for how the social world operates. After you've generated the theory, other (deductive) studies can then test hypotheses derived from it in other social contexts—hopefully, by using representative samples that will produce truly generalizable results. A common approach is for a qualitative inductive study to generate the initial theories, and a quantitative deductive study to test them. But the studies on domestic violence we just described utilized quantitative methods, which brings us back to the caveat we stated at the outset: characterizing qualitative work as inductive and quantitative work as deductive is a simplification. Sometimes quantitative researchers use exploratory and inductive approaches, and sometimes qualitative researchers use deductive approaches. The formats for qualitative and quantitative research questions we proposed earlier are just templates to get you started, and good research questions will crop up that don't fit into these cookie cutters.

Furthermore, characterizing qualitative work as always inductive—always about generating theories—can lead people to undervalue the difficult work that skilled qualitative researchers do, portraying it as simply a matter of generating ideas that "real" social scientists (quantitative researchers) go on to test. As we hope we've made clear, inductive research is important in its own right. We can only test what we already know, and induction allows us to be surprised by what we observe in the real world. Qualitative work is especially useful when we are confronted with new phenomena and complex causal relationships—when we still need to tease out how one factor affects another, how processes build upon each other, and how feedback loops and reverse causation may be influencing our results.

All that said, qualitative sociologists can shift into a more deductive mode of analysis when they are gathering data in the field. Even within the same study, they can move between modes of *discovery* and *validation* (Wilson and Chaddha 2009). When they are in a "discovery" mode, they are identifying patterns and crafting new theories in the inductive fashion we've previously described. When they are in "validation" mode, however, they look for any cases that contradict the theories they have generated so far—those that don't fit the patterns they've earlier identified. They then try to explain these outliers, refining the theory at hand to account for them. This latter strategy doesn't give us as much confidence about the existence or nonexistence of causal relationships as well-done deductive studies—with their representative samples—would, but it can tell us about contexts in which a phenomenon does *not* apply, and help us understand why. In doing so, it fits with the advice of the sociologist Charles Ragin (2023), who argues that there are no real "negative"

cases—instances that don't fit into your theory—but only "positive" cases that have yet to be matched with the particular theory that explains them.

We noted earlier the tradition of grounded theory in ethnography, which takes a purely inductive approach: ethnographers enter the field with no preconceived notions and knit together new theories fashioned solely from the data they collect. There is something to be said for this open-minded orientation toward fieldwork, which can allow us to arrive at highly original and unconventional theories. However, the truth is that even qualitative researchers who favor inductive investigation rarely take such an extreme stance. More often than not, they bring certain theories into the field, which help them figure out what sorts of questions to ask or what sorts of observations to conduct. In fact, some qualitative researchers favor a hybrid **abductive approach**. In this type of empirical investigation (to simplify), researchers apply a particular theory to the social context they're examining and then look for outliers, or deviations, from that theory. (Here, qualitative researchers are pushing hard on the "validation" strategy that we just described.) They then attempt to explain those outliers, refining the theory and making it more complex and comprehensive. Searching for empirical surprises in this way is therefore not just about generating new theory, but also about testing—to the extent possible with the limited sample at hand—whatever previous theorizing has proposed.

Key Takeaways

1. Whether a researcher takes an inductive or deductive approach will determine the process by which they attempt to answer their research question.
2. The inductive approach involves beginning with a set of empirical observations, seeking patterns in those observations, and then theorizing about those patterns.
3. The deductive approach involves beginning with a theory, developing hypotheses from that theory, and then collecting and analyzing data to test those hypotheses.
4. Inductive and deductive approaches to research complement one another and can be employed together—even, at times, in the same study—for a more complete understanding of the topic that a researcher is studying.

4.4. Refining Your Research Question

Learning Objectives

1. Identify and explain the key features of a good research question.
2. Explain why it is important for sociologists to be focused when creating a research question.

Now that you've posed a working research question in a deductive or inductive fashion, what happens next? First, you need to read as much as possible about your topic. The processing you do of past research is called a **literature review**, which amounts to a summary and analysis of the relevant publications on a topic. In **Chapter 5: Research Design**, we describe strategies for you to get the most out of the existing studies on your topic. As we describe, the literature review is essentially an argument for your research problem and the research question (and proposed research study) that addresses it.

Your working research question will evolve as you engage with the literature on your topic. You will learn new and important concepts that may shift your focus or clarify your original ideas. In fact, you should see the creation of a good research question not as a single moment of brainstorming, but as an *iterative* process: you create a working question, then do some reading of the existing literature, then refine your question, then do some more reading—rinse and repeat until you're good. Trust that a strong research question will emerge from this process at the end.

Once you figure out where the gaps in the existing literature lie, and what sorts of contributions your own research can make, you should revisit your research question and see what needs to be tweaked. Make sure to spend adequate time rethinking and refocusing your question. After all, the overall contribution of a research study often depends on the specific question that it asked. Even if you execute the study perfectly—with thoughtful interviewing, precise polling, sophisticated data analysis—your research may turn out to be entirely forgettable if your question was poorly chosen.

To conclude this chapter, let's go over some pointers for crafting good research questions that are empirical, correctly structured, and feasible to answer:

1. **A good research question is actually a question.** You can't say that your research question is "the opioid epidemic" or "oppression." You need to frame your topic as a question, not just a subject area. As we noted before, this doesn't mean you literally need to write your research question with a question mark. (Using rhetorical questions generally doesn't come across well in academic writing.) For instance, rather than stating, "My research question is: to what extent do people find sponge creatures to be morally reprehensible?" you could write, "In this paper, I will examine the extent to which people find sponge creatures to be morally reprehensible."

2. **A good research question is well-focused and clear.** You won't be able to answer everything everywhere all at once, and you should be kind to yourself, as a novice researcher, and not tackle more than you can handle. In addition to striking out any irrelevant information in your question, try your best to limit the number of concepts you mention. Break down broad concepts into their constituent parts and see if you are truly interested in just one of those parts. Of course, you may ultimately decide to keep a wide net in your data collection so that you can understand diverse, complex, and intertwined factors, which is a particular strength of qualitative studies. Even then, however, you should be able to prioritize to some extent—noting particular causal mechanisms or subjective viewpoints you intend to explore in your data collection, for instance, or at the very least having specific examples and categories in mind of what you might investigate. The definitions of any concepts mentioned in your research question should be clear, too. Those definitions do not necessarily have to be part of the research question itself, but they should appear somewhere in your write-up if they aren't obvious by context.
3. **A good research question cannot be answered with a simple yes or no.** Let's say you're conducting a study of the relationships between a married person's gender and what household tasks they do after work. Although you might have something specific in mind when you say "household tasks," your respondents might have their own impression of what the phrase means. For instance, they might think that managing finances and performing home maintenance are household tasks, but you might be primarily interested in other kinds of tasks, like childcare and cleaning. Therefore, you'd probably want to say upfront what you mean by the term (perhaps in a sentence following the research question) and also discuss why you defined it in that particular way, based on your own reasoning or arguments made in the literature. Another pertinent consideration with this example study would be how broad or narrow to go. For instance, you might want to focus your research question by picking a specific household task, or category of tasks, that you find personally interesting or that the literature indicates is important—for example, childcare. Alternatively, you could justify having a broader view of household tasks—one that encompasses childcare, food preparation, financial management, home repair, and care for relatives. In general, you should opt to keep your research question tightly focused—most students pose questions that are way too broad—but that decision ultimately rests on the specific phenomenon you're interested in and the specific research problem you want to address. For instance, your study might be more manageable if you just look at differences by gender in the efforts that married couples put into childcare. If, however, you ultimately want to understand how the unequal division of childcare responsibilities fits into broader inequalities in household labor, then you wouldn't want to narrow your question to look at just one category of tasks.



Make sure that all the terms in your research question are clearly defined. For instance, if you're studying how gender affects the ways that married couples divide up household tasks, you want to be very specific about what you mean by "household tasks." Are you talking about childcare, or all the work that married couples do in and around the home? Are you examining just physical labor or also the cognitive labor of planning and scheduling? *William Fortunato, via Pexels*

We appreciate that it can be challenging to be specific when you are just starting out your reading on the topic. (And again, that's generally the best way to develop your research question—read as much as you can.) Very broad research questions like, "What are the causes of chronic homelessness, and what can be done to prevent it?" are common at the beginning stages of a project. However, as you examine more of the literature on the topic, you'll gain the background you need to make your question more focused and clearer, and you should push yourself continually in that direction. After all, a study with a poorly posed research question is likely to get derailed at some point, given how much harder it is to creatively build upon existing research and effectively employ research methods in the pursuit of an unwieldy and vague question.

To help you focus and clarify your research question, **Table 4.3** lists a few words to watch for when evaluating it. You should avoid these and other vague terms that do not say much; they are clues that your question is not as tightly written as it could be.

Table 4.3. Words to Watch for in Research Questions

Words to Watch For	How to Be More Specific
factors, causes, effects, outcomes	What causes or effects are you specifically interested in? What causes and effects are more important, according to the literature on your topic? What causes and effects have been understudied? Try to specify one or a handful to focus on, or home in on a specific type or category of causes or effects.
effective, useful, efficient	Effective at doing what? Effectiveness, utility, and efficiency are vague terms. What specific outcome are you expecting?
etc., and so forth	Be more specific. You need to know enough about your topic to clearly address the concepts within it. Don't assume that your reader understands what you mean by "and so forth."

4. **A good research question considers the relationships among several variables.** As we noted earlier, explanatory research questions involve at least two variables—an independent variable and a dependent variable—given that they are focused on explaining causal relationships. For any explanatory study, however, you will also want to consider the role of other variables that may influence the key relationship you're interested in. Let's return to our example study on how gender affects the performance of household tasks. Even if that's the causal relationship you care about, you can't ignore the fact that a person's preferences in this area are shaped by their political and cultural perspectives. For instance, more traditional and socially conservative individuals in heterosexual relationships may be more likely to see household tasks as "women's work" and may be less likely to expect their male partners to contribute to those tasks. Because gender doesn't exist in a vacuum, your analysis will need to consider characteristics like political orientation that work together with gender to shape people's behaviors and preferences. You might therefore refocus your research question on how gender *and* political orientation affect the performance of household tasks. Note that if you are primarily interested in the independent variable of gender, you won't necessarily want to give political orientation—a variable of secondary interest—equal weight within your research question in the way we've just done. As noted in **Chapter 3: The Role of Theory in Research**, you might instead use the language of **control variables**—variables of secondary importance that still need to be accounted for in the analysis—and phrase your research question as something like, "How does gender affect the performance of household tasks, controlling for political orientation?" Alternatively, you might mention moderating variables or **mediating variables**: "How does political orientation moderate (or mediate) the relationship between gender and the performance of household tasks?" Of course, you don't want to throw too many variables into your question—remember, a good research question is well-focused!—and it's ultimately for you to decide whether and how to connect your study's variables within your research question. That said, if you

choose not to mention these secondary factors in the research question itself, you'll probably want to discuss them somewhere early on in your eventual write-up of your study to ensure that your readers know that you've considered them.

5. **A good research question identifies a target population.** To reiterate, your study's target population (or population of interest) is the larger group of people your study is interested in learning more about. It is different from the sample, the portion of that target population that you can observe and that stands in for the population you're really interested in (we cover these terms in depth in **Chapter 6: Sampling**). Your research question should be clear about the population it is targeting. After all, a study of the effect of gender on the performance of household tasks would look very different if we conducted it on a sample of older adults or a sample of newly arrived immigrants. The act of conducting research always implies that you have a target population in mind, and you should think carefully about this population as you craft your research question. Usually, you should decide your target population based on the existing literature: what groups have been adequately studied, and what groups have been relatively ignored by researchers? One easy way to create a research question is to apply a question asked by previous scholars to a new population, to see if the same findings hold up within that different context. You may even find a single type of community compelling to study because you have reason to believe—based on the literature or personal experience—that the phenomenon you're studying looks different within that group than it does elsewhere. For instance, if you noticed issues with household tasks in your interactions with first-generation immigrants, you might want to frame your research question around that particular target population.
6. **A good research question has scientific importance.** Obviously, your research question should be sufficiently interesting to you that you are willing to put in the effort to complete the project. But it is crucial to remember that science isn't just an exercise in individual curiosity. Your question needs to be important and relevant to the scientific literature in your topic area, too. Scientific relevance can be hard to assess, and we don't want you to be intimidated by the challenge of doing something original that also builds on past work. (In the next chapter, we'll take you through the process of identifying a worthwhile research problem in the social sciences that your proposed study can feasibly address—as we think you'll see, it's tough, but not rocket science.) At the end of the day, the research your question generates should have something new to say that we don't already know. It can't just regurgitate what's already been discovered.

Although this last point is more debatable, we also want to propose that a good research question is important to the broader society. This is not necessarily something that academic researchers toiling in the ivory tower think much about, given how focused we usually are on our next conference and next publication. Of course, scientific inquiry has inherent benefits for society, which in theory can use the knowledge obtained to make people's lives better. In that very big-picture sense, any scientifically important work you do is also socially important—adding another tiny stone to the heap of knowledge that already exists. But we think you should also think about the more tangible impacts your work might have on whatever communities might read your research or be affected by the knowledge gained from it. (We discuss the different audiences for research in **Appendix A: Presenting, Writing, and Publishing**.) Research projects, obviously, do not need to address all aspects of a problem or fix all of society. But you should never forget that you as a sociologist are studying topics that by their very nature have social significance. They matter to everyday people—even if the public doesn't necessarily understand or welcome the language you use to interpret that reality, or have access to all the theories and data you as a scientist can draw upon. It is within your power as a sociologist to ask worthwhile questions and find meaningful answers.

Key Takeaways

1. A poorly posed research question can derail an otherwise well-executed study.
2. A good explanatory research question is not just a statement of a subject or topic, but it is posed as a question that considers the relationships between two or more variables.
3. Research questions should address a clear target population.
4. Your research question should have scientific and (hopefully) social importance.

4.5. References

Berk, Richard A., Alec Campbell, Ruth Klap, and Bruce Western. 1992. "The Deterrent Effect of Arrest in Incidents of Domestic Violence: A Bayesian Analysis of Four Field Experiments." *American Sociological Review* 57(5):698–708. doi: 10.2307/2095923.

Blee, Kathleen M. 2003. *Inside Organized Racism: Women in the Hate Movement*. Berkeley: University of California Press.

Blee, Kathleen M. 2008. *Women of the Klan: Racism and Gender in the 1920s*. Berkeley: University of California Press.

Doyle, Arthur Conan. 2009. *The Adventures of Sherlock Holmes*. Mineola, NY: Dover.

Fligstein, Neil. 2021. *The Banks Did It: An Anatomy of the Financial Crisis*. Cambridge, MA: Harvard University Press.

Jiménez, Tomás R. 2008. "Mexican Immigrant Replenishment and the Continuing Significance of Ethnicity and Race." *American Journal of Sociology* 113(6):1527–67. doi: 10.1086/587151.

King, Ryan, Steven Messner, and Robert Baller. 2009. "Contemporary Hate Crimes, Law Enforcement, and the Legacy of Racial Violence." *American Sociological Review* 74(2):291–315. doi: 10.1177/000312240907400207.

Kirkpatrick, David D. 2005. "Conservatives Pick Soft Target: A Cartoon Sponge." *New York Times*, January 20.

Marson, James. 2012. "Ukraine's Morality Police Probe 'Gay' SpongeBob." *Wall Street Journal*, August 15.

Milkie, Melissa A., and Catharine H. Warner. 2011. "Classroom Learning Environments and the Mental Health of First Grade Children." *Journal of Health and Social Behavior* 52(1):4–22. doi: 10.1177/0022146510394952.

Pate, Anthony M., and Edwin E. Hamilton. 1992. "Formal and Informal Deterrents to Domestic Violence: The Dade County Spouse Assault Experiment." *American Sociological Review* 57(5):691–97. doi: 10.2307/2095922.

Ragin, Charles C. 2023. *Analytic Induction for Social Research*. Oakland: University of California Press.

Rothschild, Joyce. 2021. "Participatory Democratic Organizations Everywhere: A Harbinger of Social Change?" Pp. 293–301 in *Organizational Imaginaries: Tempering Capitalism and Tending to Communities through Cooperatives and Collectivist Democracy*. Vol. 72, *Research in the Sociology of Organizations*, edited by K. K. Chen and V. T. Chen. Bingley, UK: Emerald. doi: 10.1108/S0733-558X20210000072011

Rothschild, Joyce, and J. Allen Whitt. 1986. *The Cooperative Workplace: Potentials and Dilemmas of Organizational Democracy and Participation*. New York: Cambridge University Press.

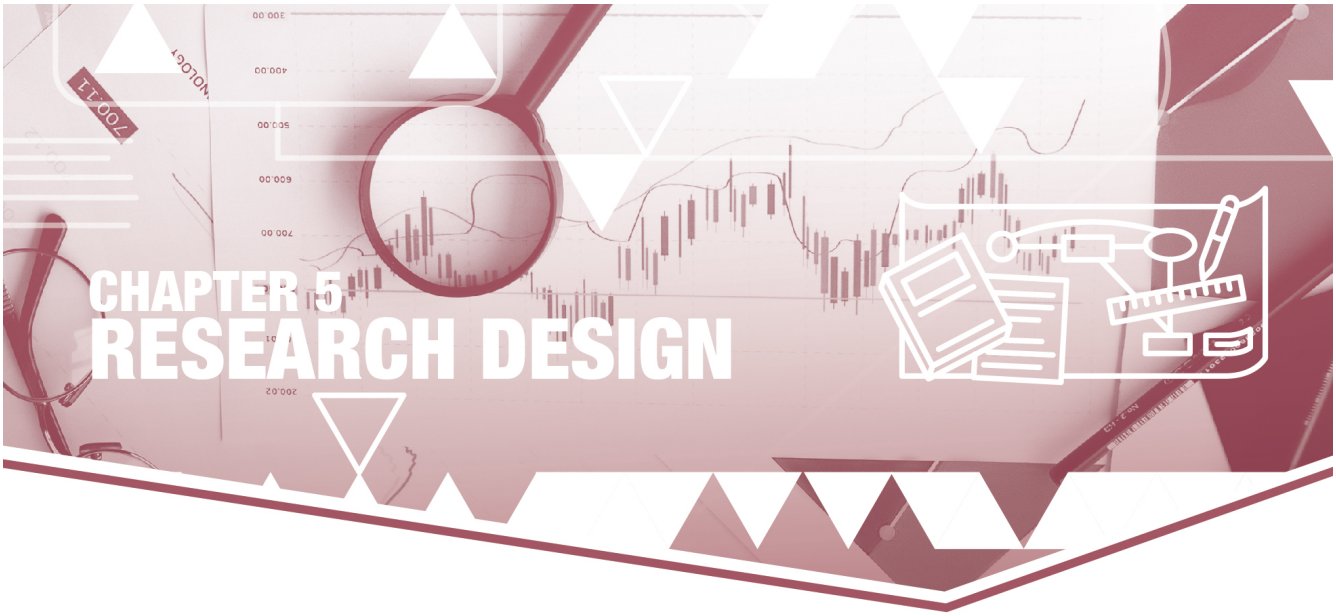
Rothschild-Whitt, Joyce. 1979. "The Collectivist Organization: An Alternative to Rational-Bureaucratic Models." *American Sociological Review* 44(4):509–27. doi: 10.2307/2094585.

Sherman, Lawrence W., and Richard A. Berk. 1984. "The Specific Deterrent Effects of Arrest for Domestic Assault." *American Sociological Review* 49(2):261–72. doi: 10.2307/2095575.

Sherman, Lawrence W., Douglas A. Smith, Janell D. Schmidt, and Dennis P. Rogan. 1992. "Crime, Punishment, and Stake in Conformity: Legal and Informal Control of Domestic Violence." *American Sociological Review* 57(5):680–90. doi: 10.2307/2095921.

Wilson, William Julius, and Anmol Chaddha. 2009. "The Role of Theory in Ethnographic Research." *Ethnography* 10(4):549–64. doi: 10.1177/1466138109347009.

5. RESEARCH DESIGN



After you have developed a working research question, the next stage in designing a research study is reviewing all the relevant literature on your topic. A **literature review** is a compilation of the most significant previously published research on a scholarly topic. It is a mandatory part of any research project. Without an adequate literature review, it is impossible to know how your proposed study can contribute to current scientific knowledge. It is quite possible that you are simply repeating what other people have done in the past—and have possibly done much better.

All sociologists must be able to look at the available literature on a topic and synthesize the relevant facts into a coherent review. Doing so shows you are deeply familiar with the body of knowledge on that topic and thereby establishes your credibility with other scholars. It also gives you the background you need to make a strong case for what the existing literature misses (this is called the **research problem**) and why your study is needed to address that gap in our scientific knowledge. Beyond their importance to making sure research actually advances science, literature reviews can have a powerful effect on policy—for example, by helping make the case that a proposed program will (or will not) have the desired impact. So even if you are not conducting original research, learning how to track down and make sense of the most relevant empirical and theoretical writings on any topic can provide useful context and ensure that you (or the policymakers you are advising) make the most informed decisions.

This chapter walks you through the steps of conducting a literature review. First, we describe what your literature review should cover. Next, we describe the sources that you should use and how to find them. Once you have identified the relevant literature, we explain how to read and evaluate your sources. Then we guide you through the process of actually writing your review. The material in this chapter is chiefly devoted to writing a literature review, given how critical this step is in determining the focus of your overall study, We also discuss how to build on your literature review and make initial decisions about the methods you will use in your proposed study. We'll conclude by discussing how to write a **research proposal.** This plan for your study includes not only a literature review that identifies your research problem, but also a methods section that describes how you plan to address that gap by collecting and analyzing actual data.

Opening chapter image credit: *Anna Nekrashevich, via Pexels. Adapted by Bizhan Khodabandeh.*

5.1. What Should My Literature Review Cover?

Learning Objectives

1. Describe what a literature review does and what forms it can take.
2. Discuss the benefits of conducting a comprehensive review of previous studies on a topic you are studying.

At its most basic level, a literature review is a survey of factual or nonfiction books, articles, and other documents published on a particular subject. That said, when you are writing a literature review for an empirical study (either one you are proposing or one you have already completed), it is better to think of your literature review as *an argument for your research problem and the research question that addresses it*. Specifically, the literature review summarizes and synthesizes what is currently known about a topic while identifying one or more gaps in that knowledge. Answering your proposed study's research question should help fill the gap that your literature review has uncovered.

Unlike the "research papers" you have written in other classes, your literature review will not just make an argument with sources. Unlike an annotated bibliography, it will not just summarize sources. Instead, it will relate those sources to your study's research question. It will make an argument for the importance of your research problem and the need for your proposed or completed study.



Remember that your literature review is not a list of sources and summaries of them. It synthesizes the literature. It identifies a gap in the literature (the research problem) and argues on behalf of a research question and study to address that gap. *raghavvidya, via Flickr*

All literature reviews should accomplish the following:

1. Introduce the topic and define the study's key terms.
2. Establish the importance of the topic (i.e., answering the "so what?" question—why exactly we should care about this phenomenon, issue, or problem).
3. Propose one or more research questions to be examined, and provide an overview of the relevant scientific literature on any concepts covered by that question, along with any related concepts.
4. Identify aspects of a phenomenon where there is not enough research and where scientists lack a full understanding (i.e., the topic is "understudied").
5. Point out areas of scholarly agreement where studies have consistently arrived at the same conclusions, as well as areas of controversy where no such consensus exists.
6. Arrive at a synthesis that organizes and categorizes what is known about a topic—not just summarizing related studies, but prioritizing the studies most relevant to the research question and highlighting the key takeaways and omissions in each related body of research.
7. Discuss the implications and limitations of previous studies and the directions that future research might take to address any gaps in that literature. When conducting a literature review for a new research project, this will entail justifying the current study—flagging one or more research problems and stating how the proposed or completed study will resolve them.

There are many different types of literature reviews, including those that focus solely on methodology, those that are more conceptual, and those that are more exploratory. When they are part of a write-up of original empirical research, literature reviews usually appear as a section near the beginning of the proposal or paper. They are meant to justify the proposed or completed study, and the literature review section will be followed by a discussion of the new study's methods (and for completed studies, its findings). Literature reviews can also be stand-alone documents. To provide other scientists with an overview of a noteworthy area of research, scholars will write long-form literature reviews called **review articles**, which may cover fairly broad topics of scientific interest (such as the review articles that the *Annual Review of Sociology* routinely publishes, which discuss the state-of-the-art sociological research on a given subject). To inform policymakers about the relevant scientific knowledge on a pressing policy area, social scientists sometimes write **policy briefs** that break down the findings of previous research, often in layperson's terms.

Regardless of their goal and audience, strong literature reviews have similar characteristics. They should be original works based on an extensive critical examination and synthesis of the relevant existing research on a topic. They should be reasonably complete summaries of the current state of scholarly understanding—that is, not restricted to a few journals, a few years, or just one methodology. And they should be arranged by key themes or findings, concluding with a discussion of areas where research is needed.

The scientific impact that your proposed research study will have depends on what others have found before you. In your literature review, try to place the study you wish to do in the context of previous research and ask, "Is this contributing something new?" and "Am I addressing a gap in knowledge or a controversy in the literature?" If you review the existing literature well, you will learn whether your initial research questions have already been addressed in the literature, and whether there are newer or more interesting research questions available. In this way, your literature review provides assurances that you are not repeating a study done many times before and that you are not making an obvious error that others have already encountered.

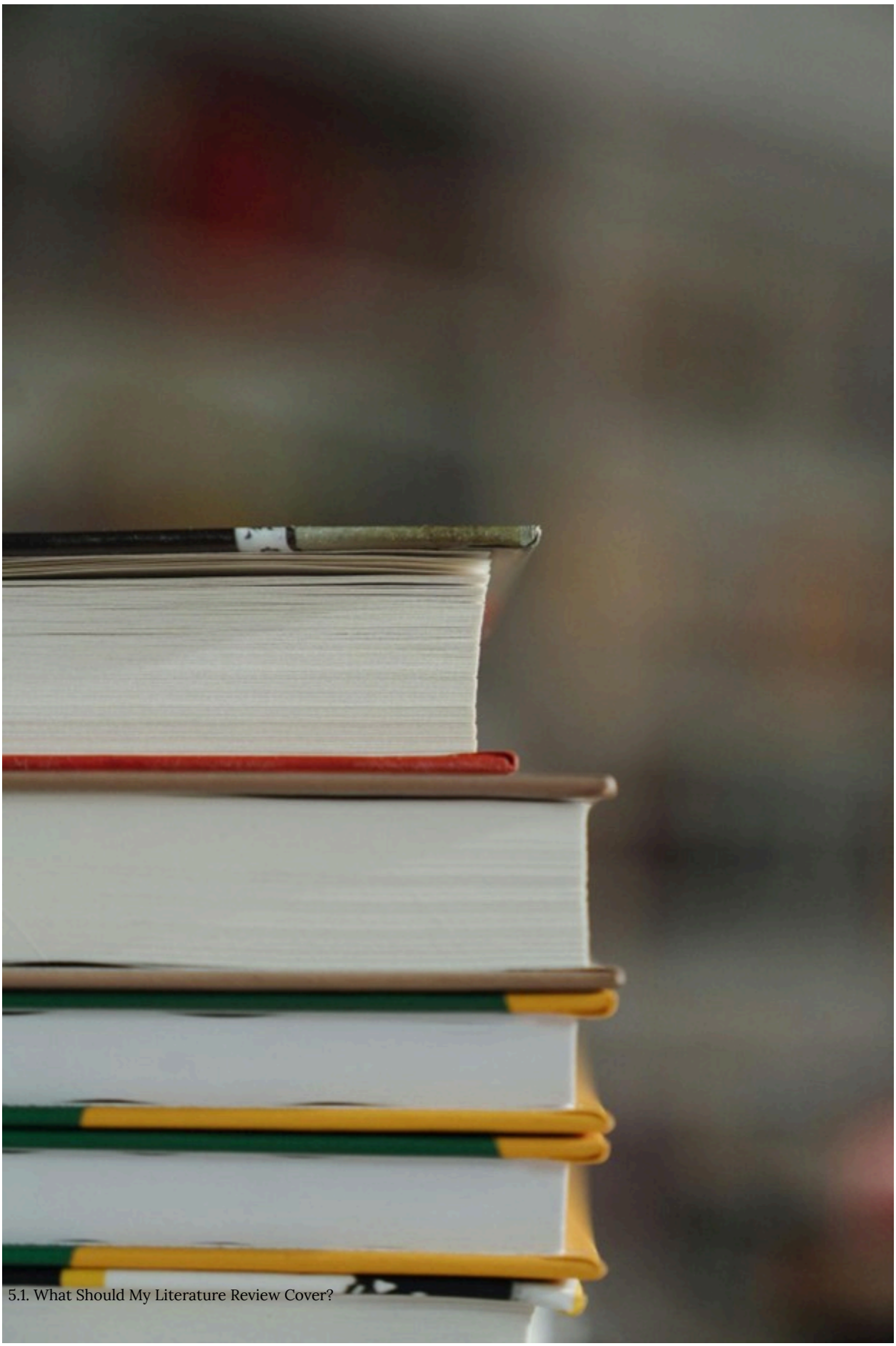
Sometimes, you are studying a topic that is so new or obscure that nothing seems to be written about it. Avoid writing a literature review that simply says, "Not much here." First of all, make sure you have found every single academic source on the topic, following the advice provided later in the chapter on doing targeted searches of the literature. If there is no sociological research on your topic, search for studies in other fields. Second, consider **parallel literatures**. These are bodies of research that do not directly address your particular research question but touch upon some relevant aspect

of it. Think of these studies as being situated parallel to an important angle of the research you want to do—for instance, elaborating on a key concept you can use in your own work. By diving into a parallel literature, you might uncover theoretical or empirical perspectives that help you answer your research question more efficiently or creatively.

Let's say you are studying how sporting events create a sense of collective meaning. In this case, the existing sociological research on rituals would be a relevant parallel literature even if that body of work currently includes few or no studies on how rituals are used in sports. Among other things, this line of research could give you a useful theoretical framework—say, to help you understand how rituals like national anthems and after-match handshakes might operate to generate solidarity among athletes.

As we noted in **Chapter 4: Research Questions**, you will undoubtedly revise and refocus your working research questions in light of what you find during your review. Drafting the literature review, much like reworking your research question, is an iterative process—one that stops, starts, and loops back multiple times before completion. You will often find yourself reading more as you work through your literature review, and stopping your reading to make sense of what you've just read by writing it up. Resist the urge to try to read "everything" and then draft your review in one sitting; instead, give yourself permission to shift between the two modes of reading and writing, which build on each other.

By reading what others have argued and found in their work, you become familiar with how people talk about and understand your topic. You encounter theories that have previously been used to address similar research questions. And you piece together possible answers to your specific questions, which will become your study's hypotheses in your later research proposal.



A literature review is not an *annotated bibliography*, which lists multiple sources and summarizes each book or article, one after another. Instead, a literature review synthesizes the findings of studies related to your research question, categorizing them by useful themes and identifying the key patterns (and exceptions) to be found within this literature. When you are writing a literature review to argue on behalf of the importance of your own work, do not give each source equal weight; reserve extended and detailed discussions of methodology and findings for those studies that are most similar or relevant to your own. *cottonbro studio, via Pexels*

As you begin your literature review, try to avoid the following common errors:

1. **Dedicating insufficient time to literature searching.** To be a good scholar, you need to be a good reader of scholarship. Intensively reading the available literature will gradually give you a sense of where the existing research falls short, and what interesting questions still need to be asked. Even if there doesn't seem to be much work on your topic just yet, don't give into intellectual laziness. Read parallel literatures and figure out how they might help you approach your research question in a novel way.
2. **Accepting another researcher's finding as valid without evaluating their methodology and data.** Based on your training as a researcher, you can read other scholars' papers and determine for yourself whether the data and methods they used were adequate. You can also assess the validity of a particular study's findings by comparing them with what others have found; in general, your literature review should prioritize results obtained from superior data and methods.
3. **Ignoring contrary findings and alternative interpretations.** In a literature review, you should describe the degree of scientific consensus regarding the phenomenon of interest, and you should make sure to point out any disagreement. Likewise, you should discuss all the possible explanations for a phenomenon—not just the one you favor—and assess the evidence in an evenhanded manner.
4. **Reporting isolated results rather than synthesizing them.** As opposed to just noting findings, one study after another, try to categorize them and group together similar studies in your discussion. Then identify any consistent findings in the literature, along with any exceptions and disagreements. Try to present those findings in a hierarchical order, highlighting the key points and subpoints—perhaps broken up into separate sections of the literature review, and then in separate paragraphs in each section. For studies that are more pertinent to your research question, you should describe both their findings and their methods at greater length.
5. **Relying too heavily on secondary sources.** As we discuss later, you can use review articles and other summaries or commentary on previous empirical studies (*secondary sources*) to help guide you in your search of the literature, but you should not lean too heavily on them. You need to read the actual empirical studies, which will sharpen your understanding of where your own study can contribute.
6. **Overusing quotations from sources.** You only need to use quotations in a literature review when you want to pass along a particularly consequential or eloquently written statement from another author. Otherwise, just paraphrase. Summarizing other people's work in your own words makes your writing livelier and has the added benefit of helping you internalize that information.
7. **Using findings that are not clearly related to your own study or using findings that are too general.** As we will discuss, your discussion of the literature should be focused enough that you are not talking in minute detail about studies that are not closely related to your own.
8. **Not justifying arguments using specific facts or theories from the literature.** In a literature review (and in any part of an academic paper, for that matter), you need to back up any claims that are not universally accepted. If other scholars have already provided theories or evidence of these claims, you should summarize those theories and empirical findings and cite the authors. If a theory is not adequately supported with empirical evidence, you should make a note of that. You also should be transparent about any lack of evidence more generally, qualifying your statements in the literature review to convey, for example, that a particular claim "may" be true, or is "suggested" by the limited data available.

Key Takeaways

1. Literature reviews are the first step in any research project, as they help you learn about the topic you chose to study and demonstrate that you can find and understand previous work in the area.
2. You must do more than summarize sources for a literature review. It is an argument for the research problem and research question: it identifies gaps in the existing literature and describes how your study addresses those gaps.
3. A good literature review puts your research in an intellectual context and helps you avoid just redoing what others have done before you. You will learn who else is studying the same topic and how you might be able to build on their work. To that end, your review must acknowledge opposing viewpoints and controversies within the literature.
4. In addition to helping you design a specific study, literature reviews make you a better scholar more generally, by increasing the breadth and depth of your knowledge of a particular topic and giving you the opportunity to read the seminal works in your field.

5.2. What Sources Should I Use?

Learning Objectives

1. Describe the strengths and limitations of each type of source you might draw on for your literature review.
2. Discuss the benefits of the peer-review process and which sources it applies to.

In your literature review, you should include all the relevant published works written by researchers in your field that document a scholarly conversation on the specific phenomenon, issue, or problem you're studying. There is no hard and fast rule about what is or isn't part of the literature on a given topic, but you will usually want your literature review to prioritize studies from your own field (i.e., sociology) or subfield (e.g., economic sociology). Scientists in a particular field or subfield tend to talk to each other, and the "literature" is essentially a collection of these scholarly debates. That said, some topics can't help but be multidisciplinary—take, for instance, public health issues—and if another discipline has studied a particular phenomenon much more than your own, it behooves you to become familiar with at least some of that work, so you're not reinventing the wheel or ignoring excellent research.

Another important criterion for deciding what to include in your literature review is whether the work is peer-reviewed. **Peer review** refers to a formal process in which other researchers ensure that a work of scholarship meets the standards and expectations of the professional field. While mainstream news outlets and book publishers do have capable editors, the review process looks much different at academic journals and academic presses. Typically, peer review involves multiple reviewers who are experts on the topic at hand, and the identities of both the reviewers and authors are kept hidden, encouraging candid and unbiased reviews. In these and other ways, peer review acts as a gatekeeper (hopefully, in a positive sense), ensuring that only rigorous research is published—specifically, research that builds thoughtfully on previous work, rests on sensible strategies of data collection and analysis, and draws reasonable conclusions from the data. Although peer review is far from perfect, the process provides for stricter scrutiny of scientific publications.

In scientific fields, papers published in peer-reviewed journals account for all or most of the sources in an academic literature review. These papers fall into a few different categories. **Empirical papers** involve the collection and/or analysis of data. This can be original data that the authors gathered by conducting their own interviews, fielding their own surveys, and so on. An empirical paper could also be based on data that someone else collected: in this case, researchers obtain access to that existing dataset but then conduct their own analysis of it. (We call the previously collected dataset **secondary data** and the original analysis of it a **secondary data analysis**). Empirical papers contain a discussion of theory, particularly in sections devoted to reviewing the literature and interpreting the study's empirical results, but the amount of theorizing can vary greatly. In contrast, **theoretical papers** (or *theory papers*) focus entirely on building, refining, or challenging theories. The authors will likely comment on empirical work—perhaps to show support for one of their theories or to challenge the validity of previous theories—but there is no original data collection or analysis.

Review articles are another notable kind of paper that you will come across in academic journals, especially in journals like the *Annual Review of Sociology* and *Sociology Compass* that specialize in publishing these articles. They are essentially long-form literature reviews not tied to a single empirical study. In addition to surveying the literature on a

broad topic, they typically make some overarching arguments about how previous empirical and theoretical work fits together, how best to understand a phenomenon or issue based on existing scientific evidence, and what important omissions exist in that body of research. To the extent they make such arguments, review articles are a kind of theoretical paper, though they can vary considerably in how much they fashion new theories as opposed to describing and categorizing past work. In empirical papers, the literature review often ends with authors making the case for the importance of their own study (building on the previous scholarship just discussed), whereas review articles do not have a specific study to justify because they don't contain original data collection or analysis. They do, however, frequently put forward a *research agenda*—that is, a list of unresolved research questions and suggestions for the types of studies that could address critical gaps in the literature.

In your search of the literature, you may also come across *scoping reviews* and **systematic reviews**. (These terms are more frequently used outside the social sciences.) A scoping review is fairly similar to a review article in a sociology journal: here, an author is trying to determine the scope or coverage of a particular topic in the current body of relevant research (Munn et al. 2018). To that end, scoping reviews will broadly describe key concepts, theories, and evidence and identify any gaps in the literature. Systematic reviews synthesize past research, but they tend to be focused on a narrow empirical question. More common in medical and policy fields, systematic reviews typically attempt to make precise conclusions about the effect of a specific intervention. They usually draw on the existing quantitative literature, although some cover qualitative work. A common type of systematic review, a **meta-analysis** (also called a *meta-synthesis*) uses statistical procedures to synthesize the results of many quantitative studies. It is basically a quantitative empirical study that treats past findings as secondary data, calculating the average effect of an independent variable on a dependent variable across studies that measure the same or similar relationships.

When writing a research proposal for your own empirical study, you should incorporate a mix of theoretical and empirical papers in your literature review. Review articles and meta-analyses can be helpful to you in finding or framing the relevant literature, but they won't show up in every literature review. Adequate coverage of the empirical work on your topic is especially important, given that your goal is to extend that prior work with new data collection and analysis.

So how do you distinguish between these different types of papers? As we discuss in **Section A.3. Writing Up Your Work for Other Scholars**, empirical papers (which include meta-analyses) follow a similar structure, with a methods section and a results or findings section somewhere in that mix. While the exact headings may differ slightly from publication to publication, the presence of these two sections will signal to you that you're reading an empirical paper. Theoretical papers and review articles will not include methods or findings sections. In fact, these papers do not follow a set structure. Authors can more or less organize their arguments in whatever way is convenient. We should also emphasize that just because a paper *mentions* quantitative or qualitative data does not mean it is an empirical paper. Most papers contain literature reviews, and most of those reviews discuss empirical findings of some sort. For a paper to actually be an empirical study, it needs to report findings from the author's own analysis.

Let's talk more about the types of sources that might be included in your literature review. We'll talk first about papers published as articles in peer-reviewed journals, which will typically be the majority of sources that you cite. In sociology, academic books are also commonly included in literature reviews. Other sources can be helpful, too, but as we will describe, you should be more cautious about relying on them.

Journals and Journal Articles



Nowadays, you often don't need to step foot in a library to read the state-of-the-art literature on social scientific topics, which instead will appear in digitally published academic journals. University libraries pay exorbitant fees for their faculty and students to be able to access the online databases that contain these papers. Note that it is scholarly etiquette to give a copy of any paper you publish to other researchers who request it, so you can politely request an article from its author if it is hard to obtain otherwise (asking for published book manuscripts is less common, however). *Tima Miroshnichenko, via Pexels*

Papers are typically published in **academic journals**, periodical publications that contain a number of papers that reflect the journal's focus. Academic journals can be **generalist journals** that publish any noteworthy paper in sociology or another discipline, or they can be **specialist journals** dedicated to a subfield within the discipline (like the sociology of education or organizational sociology). The mission of any journal is to publish work that advances knowledge about the field or subfield.

Journal articles are written by scholars with the intended audience of other scholars interested in the subject matter. They contain extensive references to other studies in the relevant literature, so that readers can better judge the arguments and data and so that the authors connect their research to the work of other scholars in the same area. Professional sociologists often read one or more generalist sociology journals regularly, along with one or more specialty journals in their subfield. As a result, publishing your paper in a sociology journal is often the best way to reach other scholars who are interested in the topics you study.

Scientific journals almost always vet the papers they publish through peer review. Their editors recruit experts in each paper's area of research who confidentially assess its merits and advise the journal on whether to publish it. For their paper to be accepted, authors usually must revise their work substantially in response to any reviewer feedback, which can go on for multiple rounds. (For more details, see the sidebar **Peer Review: What Makes a Science a Science**.) Because of this rigorous process of evaluation and revision, journal articles are the main source of scientific information that most researchers rely upon, and they dominate any literature reviews written on a particular scientific topic.

You can often tell whether a journal is peer-reviewed by going to its website. Usually, the website will describe its procedures for peer review. You can also search on a university library catalog; peer-reviewed journals are typically listed as such in their catalog entries. Finally, you can turn to periodical databases such as Ulrichsweb that record whether a particular publication is peer-reviewed.

Note that journals have different reputations for the rigor of their peer-reviewing process and the quality of the work they publish. The "top" journals in the field—for U.S. sociologists publishing empirical papers, this would include the *American Sociological Review*, *American Journal of Sociology*, and *Social Forces*—are able to tap nationally or internationally recognized experts as reviewers on virtually any substantive or methodological topic. These publications also reject the vast majority of submissions. Other journals do not have such strict standards. That does not mean they publish shoddy research, but you want to be more cautious with less selective journals, knowing the review process may not have been as intensive. You should also be aware of so-called *predatory journals*, which are meant to generate money for the publishing company by charging scholars to have their work published. (Given the difficulty of finding reviewers, some prestigious journals have submission fees, but those fees tend to be small—serving to discourage less serious submissions.) Predatory journals may claim to be peer-reviewed, but in reality, they publish any article—no matter how poor it is—as long as the author pays them.

How do you know which journals are reputable? Perhaps the best way is to read a lot of the literature in your research area and see which journals keep cropping up in the reference lists. These are likely to be reputable journals (since scholars are citing them) and may even specialize in topics like yours, meaning you'll want to be a regular reader of that journal. You can also get a sense of a journal's reputation in the field by looking up its **impact factor**. Databases like Journal Citation Reports rank academic journals based on how many citations their articles received a year or two after publication—a rough measure of how influential those articles are, and therefore how respected the journal

that publishes them is. One of the world's most prestigious academic journals, *Nature* (which publishes articles across scientific disciplines), had an impact factor of 64.8 in 2022, according to Journal Citations Reports. This means that the papers that were published in *Nature* in 2020 received, on average, 64.8 citations in 2022. Journals in sociology—a smaller scientific field without the sizable practitioner communities of the hard sciences—tend to have much lower impact factors. And just because a journal does not have a high impact factor does not necessarily mean it is a disreputable journal: journals that specialize in niche subjects may have particularly low impact factors, and yet the most prominent scholars in a subfield may be contributing to them. All that said, journal impact factors provide another way that you can check the trustworthiness of a source you're considering using.

Part of the tuition your university or college charges you goes to paying the publishers of academic journals for the privilege of accessing their articles. Because access fees are incredibly expensive, your school likely does not pay for access to all the journals in the world. Nevertheless, you should never have to pay for access to an academic journal article while you are in school. If your library does not subscribe to a journal you need to read, you can usually go through its website to order the article through interlibrary loan (you can follow the same procedure to get a copy of an academic book, too). If you are not currently affiliated with a university, getting academic sources is more challenging, but you should feel comfortable asking an author for a copy of their article. Academics are not paid any royalties for journal articles (and sometimes they don't even receive them for academic books), and they will usually be happy to share their papers with anyone interested in their work. Most academic journals allow their authors to share their articles privately with interested readers, so the authors won't get in copyright trouble, either.

You should keep an eye out for an important string of numbers (and sometimes letters) labelled "DOI" or "doi" somewhere on the first page of a journal article. This stands for **digital object identifier**. Based on an international standard, DOIs are unique identifiers for scholarly papers, government reports, and other documents. Most recently published scholarly materials (including this textbook) have their own DOIs. Although you should not assume by the presence of a DOI alone that an article is scholarly—much less that it is peer-reviewed—seeing one listed is a hint that you're on the right track. Note that lesser-known journals may not have assigned DOIs to their articles. Likewise, even though DOIs have been added to some older sources, for the most part articles published before 2000—when the nonprofit DOI Foundation introduced the standard—will not have them. One useful feature of DOIs is that you can enter them into a web browser to be sent directly to the official webpage for that document, such as the article's page on its journal's website. Simply copy and paste the DOI after **https://doi.org/** in your browser. For instance, if the article's DOI is 10.1177/15365042221131082, then https://doi.org/10.1177/15365042221131082 will take you to that article's webpage.

Peer Review: What Makes a Science a Science

The peer-review process has several stages. First, after a scholar submits a manuscript to a peer-reviewed journal (also called a **refereed journal**), the journal's editor (or editorial team) will evaluate it. They may opt to reject the paper outright (a *desk* rejection) or "send it out for review." In the latter case, the editor will reach out to scholars they know of—through their own reading or through their professional networks—and recruit two to four people (sometimes more) as peer reviewers. Ideally, the editor will gather a group of reviewers who complement one another and cover all the areas of expertise required to properly assess a particular manuscript—say, one reviewer who is knowledgeable about the substantive topics being discussed, along with another who is well-versed in the methods being used.

Once recruited, these reviewers—the "peers" in "peer review"—read the submission and help the editors to decide whether to publish it in the journal. Sometimes, they rate and evaluate the paper in confidential feedback that only the editors read. The reviewers' primary responsibility, however, is to write detailed

comments about the strengths and weaknesses of the manuscript that are then shared with the authors. The feedback sent to the authors will not have the names of the reviewers attached, and typically the reviewers will not know who the authors are, either. We call this process of keeping identities secret on both sides **double-blinding**: the reviewers are "blind" to the identities of the authors, and vice versa. To ensure that the reviewers cannot guess who the authors are, editors will ask authors to "blind" their manuscript by removing any identifying material, such as mentions of their past publications.

Scientists see double-blinding as essential to ensuring that peer reviewers provide honest and impartial feedback. If reviewers knew the identity of the authors, their personal feelings might consciously or unconsciously bias their evaluation of the manuscript: they might be predisposed to believe that any paper written by a high-profile sociologist will be genius work, for instance, or they might be loath to recommend accepting a paper written by someone they've publicly quarreled with or privately dislike. Likewise, if the authors know the identities of the reviewers, the reviewers are likely to be less candid in their feedback—they might pull back on any criticism, for example, because they fear a backlash from the authors.

Once the peer reviews are completed, the publication's editor (and sometimes its editorial board as well) read through the reviews and decide whether the manuscript provides a noteworthy contribution to the field and should be published. After the first round of reviews, it is highly unlikely that an editor will outright accept a manuscript, even if it is exceedingly well-done. Instead, they will either reject it or ask for revisions based on the reviewers' (and possibly their own) feedback. The latter decision is called "Revise and Resubmit," and it is typically a good outcome for authors to receive: they are still on the path to publication.



Having anonymous peer reviewers review a study before it can be published is a central feature of modern science, and this vetting process helps ensure that only rigorous and thoughtful research gets published. However, the anonymity that reviewers enjoy can mean that cranky individuals with axes to grind can derail an otherwise promising study—as epitomized in memes devoted to mean-spirited "Reviewer #2." *xixoxixa, via Reddit*

When they resubmit their manuscripts, the authors also need to write a letter (a *response memo*) documenting whether and how they responded to each of the points made by their reviewers. If they chose not to address a certain suggestion, they need to justify that decision; if they did make revisions in response, they need to describe them. The editor will review the revisions and response memo and decide whether to send the manuscript out for another round of review—sometimes with the same reviewers, but sometimes with one or more new reviewers. Depending on the journal, this process can continue for additional rounds of revision—with new reviews and new response memos each time—until either the manuscript is rejected or accepted. Usually, editors will reject a manuscript if the peer reviews are negative or they don't feel the authors adequately addressed the reviewers' feedback. They occasionally intervene to accept a paper with mixed reviews when they believe strongly enough in its contribution. But even positive reviews are no guarantee of publication: as we noted, the first round of reviews almost never results in an acceptance, and at selective journals in

particular, editors may overrule the reviewers and reject the manuscript even if the last round of reviews was positive.

In sociology and many other fields, peer reviewers are generally not paid. They participate in the process because they want to stay abreast of the latest research in their field and influence scholarly conversations of a topic they care about (perhaps in the process gaining exposure for their own work). If they don't do peer reviews, they may also get a reputation as a scholar who doesn't give back to the intellectual community of their field or subfield. A journal editor they just rebuffed might look dimly on a future manuscript submission by them, too. Notwithstanding these incentives, many scholars decline invitations to review, and editors often have to reach out to a long list of potential reviewers just to find two to four. Graduate students are sometimes asked to serve as peer reviewers. Although we have concerns about the exploitative nature of the current review system, we recommend that young scholars try their hand at peer-reviewing, because doing reviews will make you a better critical thinker—especially when it comes to evaluating research methods. Learning to dissect a paper and say clearly what parts of it work and don't work will improve your research design skills and enhance the quality of your own scholarship.

Academic Books: Monographs and Edited Volumes

In some fields (including sociology), the academic literature also includes **monographs**, scholarly books that focus on a particular topic. Sometimes, chapters of a monograph will have been previously published as journal articles (which means you don't necessarily have to buy or borrow the whole book to learn the gist of the arguments being made). Monographs will often take on the structure of a journal article, too, with a chapter devoted to a literature review, a chapter devoted to describing the study's methods, and separate chapters on key findings.

Monographs are often published by university presses, nonprofit publishers affiliated with universities that typically have the words "university press" in their names. University presses and other scholarly presses routinely peer-review their book manuscripts. The process is different from what takes place with refereed journals, however. For one thing, when a book author is pitching their book to a university press, that press may recruit peer reviewers to take a look at the submitted book proposal—which typically includes a few sample chapters. Those reviewers will give the press and the author their feedback on the quality of the proposal and whether the book should go forward. Later, once the book project is completed, a peer review will also be conducted on the entire manuscript (possibly by an entirely new group of reviewers). Another difference between journal and book peer-reviewing is that the reviews for book manuscripts are **single-blinded**: the reviewers know who the author or authors are (it is just too impractical to mask their identities in such a long work), although the names of the reviewers are still kept secret.

Sometimes research written by different scholars is collected together in what is called an **edited volume**—a book whose chapters are individual papers. These papers may not have made it into academic journals for whatever reason, or they may be seminal articles that are being republished in the volume. Edited volumes are considered less reputable than journal articles, as the papers collected within them are often reviewed in a less rigorous fashion (e.g., without blinding)—if they are peer-reviewed at all. That said, chapters in edited volumes are routinely cited and are considered a legitimate source to use in your literature review.

Other Scholarly Sources of Information

When writing a literature review, you should prioritize peer-reviewed journal articles and monographs as sources. Other types of material can be used, but they are generally places to start—not end—your search for sources. Consider

academic encyclopedias and textbooks. When you are first getting to know your research area, you might want to consult an academic encyclopedia like the *Blackwell Encyclopedia of Sociology*, which covers a wide range of sociological topics and theorists, or the *Stanford Encyclopedia of Philosophy*, which has briefs on a number of social theorists. Some textbooks (including this one) are peer-reviewed by experts, and they can include useful discussions of, and citations for, literature relevant to your study. However, professional sociologists rarely mention these works in their papers. Instead, they will go straight to the sources that these sources cite and branch out from there. Review articles do tend to be cited in professional scientific writings—especially if they have a particular interesting theoretical take on the literature, or if the person citing the review article wants to briefly refer to a body of literature that is not central to their research question (e.g., "for more on this topic, see this review article"). But again, it is usually best to go straight to the original sources that the article mentions.

Another kind of scholarly source is **gray literature**, which is research and information produced by nonacademics. These can be reports by government agencies like the Centers for Disease Control and Prevention or U.S. Census Bureau, write-ups of survey results by polling outfits like Gallup, or policy briefs by think tanks like the (progressive) Center on Budget and Policy Priorities and (conservative) Manhattan Institute. As we mentioned in **Chapter 2: Using Sociology in Everyday Life**, government agencies, pollsters, and think tanks have researchers on staff who are trained social scientists. As a result, you may find yourself drawing upon their analyses and interpretations of data for your literature review, particularly when you want to describe a particular context. (The official statistics that government agencies provide are invaluable in this regard.) That said, this kind of research is not peer-reviewed and rarely engages with theory, which makes it less useful to sociologists. It also frequently espouses a specific political viewpoint in line with the organization's mission, which may make any findings suspect.

At times, you will come across writings by scholars that are published in nonacademic venues. For instance, academics occasionally write books for mainstream commercial presses (called **trade presses**) rather than university presses. Similarly, scholars can write about their research or related issues in articles published by news outlets. The writings they produce for these outlets frequently take the form of commentary pieces, or **op-eds** (we discuss how to pitch and write op-eds in **Appendix A: Presenting, Writing, and Publishing**). Sociologists may also write about their research on personal or organizational blogs; various sections of the American Sociological Association host their own blogs, and some blogs written by social scientists have become influential in wider policy debates.

Trade books and op-eds are not peer-reviewed. That said, they may be subject to intensive editing and even fact-checking, depending on the publisher. They also may engage in scholarly debates and discuss theories and empirical findings that might be useful for literature reviews. As a result, these works are sometimes heavily cited in academic circles, although really their audience is the general public (or whatever audience the publication or publisher caters to). You can use them as sources, but do not rely on them too heavily; instead, go to the journal articles or monographs that are usually the basis for the scholar's popular writings.

You may come across a variety of academic "works in progress" that can be included in your literature review, but you generally will not want to use them if more polished work is available. For instance, dissertations and theses—which often go on to be published as journal articles or monographs—can be rich sources of information and feature extensive reference lists to scan for resources. As part of the dissertation and thesis defense, a committee of professors did review the work, and it needed to gain their approval to be accepted. This is not the same thing as a peer review, however, and you may want to wait until the work is published in a journal or as a book before citing its findings. Regardless, you should definitely read the manuscript's literature review and plumb the sources the author uses for your own review.

Talks at academic conferences (which we describe in **Appendix A: Presenting, Publishing, and Writing**) are sometimes cited as academic sources. At conferences in some disciplines, the papers presented will be published in a peer-reviewed volume of "conference proceedings"—when they reach that stage, they are not works in progress but rather publications, some of which go on to be widely cited in their field. In sociology, however, conferences do not tend to impose a rigorous peer-review process for papers—and many don't even require scholars to submit completed

manuscripts in order to present their work. As a result, conference papers and presentations should be used less as sources for your literature review and more as guides leading you to other good research—for instance, other work by the same scholars on that topic, or the final version of the presented paper that is later published in a journal. If you do find a relevant conference presentation, you should ask the presenter to pass along the corresponding paper manuscript. As we noted earlier, most scholars are happy to share their work with you—whether it's a conference paper or even a published journal article.

A final suggestion about evaluating scholarly sources: when academic databases list a particular paper or book in their search results, they often include the number of other academic sources that have cited that paper or book. Like impact factors for journals, a citation count for papers or books is a blunt indicator of how influential that work of scholarship is within the research community (note, of course, that sometimes an article is cited a lot because it is controversial!). When reviewing the scholarly literature, you may want to start with well-cited books and articles so that you have covered any "must-read" studies in your research area—but remember that sociology articles in general tend to have lower citation counts than those in larger scientific fields.

Nonacademic Sources

When you start learning about your research topic, you will often come across two types of nonacademic sources. First, there may be news articles discussing the issue. These pieces are intended for a general audience and are published in periodicals (magazines and newspapers) and online news outlets. With the exception of scholarly op-eds and the like, they are usually written by journalists. Second, you may encounter trade publications—magazines, newsletters, and (nonacademic) journals geared toward people who work in certain industries or professions. Unlike magazines and newspapers, trade publications may take some specialized knowledge to understand. They often have information about industry trends and practical information for people working in the field.

Trade publications may be somewhat more reputable than general-interest periodicals because the authors tend to be specialists in their field. Although the authors' writings for a trade publication will not have gone through peer review, their take on a topic may be relevant to your literature review, especially if your research delves into practical applications. Generally speaking, however, neither news articles nor trade publications belong in your literature review. As we have emphasized, your proposed study needs to be connected to the scholarly conversation in the scientific literature, and reading too many nonacademic sources can be distracting and divert you from that conversation.

Nonacademic sources do have some uses in academic writing. For one thing, news articles are frequently used as data in a **content analysis**, which is a scholarly analysis of existing text and media often intended to explore popular views of an issue within a particular context (we will cover content analysis in **Chapter 15: Materials-Based Methods**). For a literature review, nonacademic publications can provide supplementary context or help with brainstorming. They may include pertinent facts or ideas you should include, especially when you need to offer some general background on a topic (which typically goes in the introduction to a research proposal or research paper). In particular, academics rely on journalistic pieces for details about specific events or organizations they are discussing. In this case, they are drawing on the news articles as **primary sources**—first-hand accounts of an event or an organization, which they can use as data or cite in their literature reviews to back up a claim. Although these sources are not peer-reviewed, more reputable news organizations have high standards for their reporting and may utilize professional fact-checking—which sometimes means their factual accounts may be more reliable than those in academic writing. Beyond their regular reporting, news organizations at times will gather statistics that can also be of use to social scientists. For instance, news organizations like the *New York Times* and CNN regularly commission political polls. In the absence of government reporting, the *Washington Post* and other news outlets have compiled historical databases tracking police shootings.

Nonacademic sources are also a good place to start when you are doing your initial brainstorming. They can spark ideas and give you relevant and precise keywords to use in searching the academic literature, the real target of your review.

For instance, news articles often summarize the results of noteworthy scientific studies. In these cases, the articles are **secondary sources**, reporting the results of original research.

In a scholarly journal article, you will usually be given a great deal of information about the researchers' method of data collection, their sample, and the process by which they recruited their research participants. All of these details provide important contextual information that can help you assess the researchers' claims. But when you read a journalist's write-up of social scientific research in a news article, you usually won't get the same level of detail about the study. You can use these secondary accounts of research to give you a quick sense of what social scientists have found and to help you locate the relevant academic literature. However, you should not rely solely on the summaries of research in secondary sources: *you must read the original source.* Besides for the lack of detail they provide in their write-ups of studies, journalists are not scientists. You have undoubtedly seen news articles making claims about how a particular food prevents cancer or how a new treatment can extend your life, which later turn out to be wildly inaccurate or premature. Clearly, journalists can exaggerate or misinterpret results. Careful scholars will critically examine the primary source rather than rely on someone else's summary. Likewise, many news outlets also produce opinion pieces, which are not always labeled as commentary rather than news. These articles may present facts that support the author's viewpoint while excluding or downplaying facts that contradict it, and they are best avoided.

What about an organization's website, press releases, and social media feed? If an organization is a noteworthy player in the field you're studying, then looking at their posts, reports, announcements, and other online content can clue you into important topics to study. Many advocacy websites will provide references for the facts they cite, providing you with the primary source of the information. Again, though, you generally want to turn to the original source rather than citing the webpage that mentions it.

Another type of source that students go to almost immediately is general-interest encyclopedias like Wikipedia. (These are **tertiary sources**, which summarize secondary sources.) For its part, Wikipedia is a marvel of human knowledge. For the most part anyone can contribute to this digital platform, but the entries are overseen by skilled and specialized editors who volunteer their time and knowledge to ensure that their articles are correct and up to date. Nevertheless, you will almost never want to cite Wikipedia or similar general sources in a literature review. Instead, use it as a resource to get a general sense of the research on a topic and look to any secondary sources it cites.



The online encyclopedia Wikipedia and other tertiary sources are terrific places to start learning about a topic, but you usually should not rely on them as sources in a scholarly paper. Instead, use them to identify key academic articles and books to read and important topics or concepts to use as keywords in your database searches. *Wikipedia*

Just like academic sources, nonacademic sources vary greatly in terms of how carefully they are researched, written, edited, and reviewed for accuracy. Common sense will help you identify obviously questionable sources, such as highly political partisan news outlets or a random person's blog. Sometimes, however, a source's reliability is not so obvious. You will need to consider criteria such as the type of source, its intended purpose and audience, the author's qualifications, the publication's reputation, any indications of bias or hidden agendas, how current the source is, and the overall quality of the writing, thinking, and design.

Nonacademic sources are excellent places to start your journey into the literature, because they do not require specialized knowledge to understand the content and may inspire deeper inquiry. Nevertheless, you generally don't need to discuss or cite them unless you need attribution for a specific fact—for instance, the results of a political poll mentioned in a news article. Although nonacademic sources can be an important part of how we learn about a topic, your literature review should instead focus on finding peer-reviewed articles and books. These publications are really the basis of scientific conversations among researchers.

Key Takeaways

1. Peer review is the process by which other scholars evaluate the merits of an article before publication. Peer-reviewed journal articles and books are considered the best source of information for literature reviews, although other sources are often used as well.
2. Social scientific research requires critical evaluation of each source in a literature review. Whenever possible, go to the original study that other sources are reporting on.

5.3. How Do I Find Relevant Sources?

Learning Objectives

1. Describe useful strategies to employ when searching for literature.
2. Learn to narrow down search results to the most relevant sources.

Finding relevant sources for your literature review usually follows these steps:

1. Finding the right database
2. Building search queries
3. Skimming articles
4. Examining references

One of the joys of being a researcher in the twenty-first century is that we can do much of our work without ever leaving the comfort of our recliners. This is certainly true of familiarizing yourself with the literature. Most libraries offer incredible online search options and access to important databases of journal articles and other academic sources. (A quick note: for simplicity, we may refer to "papers" and "articles" when we mean academic sources in general.)

Rather than just diving into keyword searches on your topic, however, we recommend that you start off by doing two things. First, if you intend to draw upon an ample amount of literature—say, dozens of books and articles—use reference management software to help organize your sources. (See **Appendix A: Presenting, Writing, and Publishing** and its sidebar **Using Zotero to Manage Your References** to learn about the free and open-source application Zotero, which can be installed on your computer or mobile device and which automates some of the labor involved in collecting and citing relevant academic literature.) Second, start your search for sources by looking for review articles and meta-analyses on your topic. As we described earlier, review articles provide a summary of the existing literature on a topic. If you find one on your topic, you will be able to read one person's summary of the literature and go deeper by reading their references. Similarly, meta-analyses have long reference lists that are useful for finding additional sources on a topic. They can give you a precise sense of the scientific consensus about a particular phenomenon across however many studies were included in the meta-analysis. Remember that both review articles and meta-analyses amount to their authors' interpretations and assessments of the findings within a particular field, and there may be intense disagreements among scholars about the conclusions they draw. Nevertheless, review articles and meta-analyses are good places to start your literature search.

It goes without saying that you should not plagiarize the summaries of the literature in a review article (or any article's literature review, for that matter) when writing your literature review. (See our extended discussion of **plagiarism**, including **mosaic plagiarism**, in **Appendix A: Presenting, Writing, and Publishing**.) Instead, figure out which area of the literature you want to focus on and track down the studies mentioned there. Describe these studies in your own words and use the review articles to help you put them in some sort of context regarding their importance within the field.

Hopefully, you will find one or more overarching sources to guide your literature search. Oftentimes, though, a topic won't have a decent review article or meta-analysis to rely on—the papers you find will be rather dated, or they may discuss the topic you want to study in an overly broad or overly narrow way, or they may cover a different field than

sociology (which may not be so helpful to you, since you will want your literature review to be rooted in specific scholarly conversations that often fall along disciplinary lines). Don't fret. At some point you're going to need to move on to a general literature search, anyways. Research should always build on the most recent published studies on a particular topic, so you can't just rely on review articles or meta-analyses to ensure that you've done your due diligence. Furthermore, the scholars who summarized the past literature may have missed important work that relates to your specific research question.



Searching the expansive social scientific literature for articles just on your topic can seem overwhelming. Be smart about the databases and search queries you use to narrow down your searches and make your reading manageable. *CHUTTERSNAP, via Unsplash*

Step 1: Finding the Right Database

You can use the following major databases to find academic sources relevant to sociology:

- **Google Scholar:** This specialized search engine covers the academic literature across disciplines and gives you many options in searching for journal articles, monographs, and other scholarly sources. Google Scholar is a "natural language" search engine, which means it tries to use its knowledge of how people talk to better understand what you are looking for.
- **Sociological Abstracts:** This index provides a list (including abstracts) of the international literature from sociology and related disciplines. Sociological Abstracts is linked to Social Services Abstracts, which covers research in social work and human services.
- **SocINDEX:** This database offers the full text of peer-reviewed sociology journals covering research areas like gender studies, criminal justice, social psychology, racial studies, religion, and social work.
- **Applied Social Sciences Index and Abstracts (ASSIA):** This index contains abstracts of international journals

covering health, social services, psychology, sociology, economics, politics, race relations, and education.
- **Social Sciences Citation Index:** Part of the larger Web of Science database, this index includes articles from more than hundreds of journals across 50 social science disciplines.

As we've discussed, you want your research to be connected to existing scholarly conversations within a field or subfield. To that end, you may want to pursue a more focused search for sociology papers and books, using the databases we mentioned that are devoted to sociological sources. However, because Google Scholar is free, requires no library subscription, and is widely used, we'll mainly talk about how to find sources using this search engine, with the idea that some of these strategies can also be applied to other databases.

Note that Google Scholar by default displays not only peer-reviewed journal articles and books, but also nonrefereed articles and books—possibly with government and think tank reports and other gray literature as well, depending on whether academic publications cite them. Given the wide range of sources included in Google Scholar, you will need to take extra care in vetting the quality of the sources you find (we'll have more to say about that later in the chapter), as well as ensuring that they are focused enough on sociology or whatever subfield your research is contributing to. Those caveats aside, however, you should be able to use Google Scholar to find more than enough relevant sociology sources to start your literature review. Furthermore, we would advise you in general not to rely exclusively on searches of databases to find sources for your literature review. They are great places to start, but you should move away from them eventually and start identifying any new sources from scientific debates on your topic that are threaded across the papers you're reading—specifically, by following the web of studies that cite each other. That way, you are connecting your work to an ongoing conversation within the literature.

Note that Google Scholar can be linked to your university library login by using the Library Links option (on Google Scholar's Settings page), allowing you to access journal articles with one click on the Google Scholar page. Google Scholar also allows you to save articles in folders and provides a (somewhat correct) American Psychological Association citation for each article.

Step 2: Building Search Queries

Probably the most frustrating part about literature searching is that you will often get hundreds or thousands of articles for the searches you do. How can anyone be expected to read all these articles? First of all, you don't need to look at all the sources you find. In fact, all you really need is a few relevant articles to get you started, and then, as we discussed, you will want to follow the breadcrumb trail of citations between articles to familiarize yourself with the important scholarly conversations on your topic.

That said, when you are searching for sources using a database or search engine, you can make your life easier by choosing your search **queries**—the search terms you use—with care. A precise search query can narrow down your results to a much smaller number of articles, all of which are specifically focused on your topic area. Since Google Scholar is a natural-language search engine, it gives you the flexibility to ask questions or type sentences or phrases in your search query. However, most other databases will not return useful results unless you use **keywords**—terms relevant to your topic area. For instance, in Google Scholar you could type `the effects of cocaine addiction on the quality of parenting` and find relevant articles, but in a database you would probably want to type something like `cocaine addiction parenting child development`. (See the sidebar **Search Operators** for more tips about how to combine keywords using Boolean search operators like `AND`, `OR`, and `NOT`.) It might be even better to put the phrases `"cocaine addiction"` and `"child development"` in quotation marks, which will return only sources with those exact phrases somewhere in their text, further narrowing your search.

Including jargon used by social scientists who study this topic may also help focus your search queries and filter out off-topic articles. For example, if you wanted to learn more about people from low-income households who do not have

access to a bank account, you might use the keyword "unbanked," a term that researchers use to refer to people without bank accounts. To find papers that talk about children who take on parental roles in families, you might include the term "parentification" in your search query. Of course, as a novice researcher, you can't be expected to know these terms ahead of time. Instead, start with the keywords you already know. Once you read more about your topic, start including new keywords that will return the most relevant search results for you.

Here are some other tips for making your searches more effective:

1. **Search for your keywords in specific fields.** Many of the irrelevant results in your searches will probably be from articles that aren't on your topic but just mention your search term in passing. Fortunately, most databases allow you to search for keywords in specific fields, such as the abstract of a journal article or book chapter. If your topic isn't in the abstract, chances are the article isn't relevant. You can be even more restrictive and search for your keywords in the title field. If the database has a "subject" field or something similar, put your keywords in it to get results from authors who have tagged their articles as addressing that topic. In Google Scholar, you can search for keywords in the titles of sources using the `intitle:` tag (e.g., `intitle:"social capital"`).
2. **Narrow down the time frame of your search.** Articles older than several years won't tell you the current state of knowledge on a topic. When starting your search, it is a good idea to focus on finding recent papers. Their literature reviews will hopefully cover the latest studies on your topic in addition to discussing the seminal articles and books in your subfield. Most databases have options to narrow down your results by date. In Google Scholar's search results page, you have the option to further narrow your search to the past few years. As shown in **Figure 5.1**, use the links on the left when using a desktop browser. From Google Scholar's hamburger menu in the upper-left corner, you can also choose the Advanced Search option to narrow your search to specific dates.
3. **Search particular journals.** You'll notice that many of the articles you find in your literature searches are from the same journals. As we mentioned earlier, some journals are specialist journals that focus on particular research areas. Targeting those journals in your searches will not only help filter out irrelevant articles but also will ensure that you are engaging with the same scholarly conversation on your topic. For example, if you are searching for sociological articles related to employment and workplaces, you might want to search within the journal *Work and Occupations* and the edited volume series *Research in the Sociology of Work* for some good leads. If you are looking for the latest studies on cohabitation (couples who live together but are not married), you might want to search just the *Journal of Marriage and Family*. In the latter case, you could use Google Scholar's Advanced Search option and enter `"journal of marriage and family"` (note the quotation marks to search for an exact phrasing) in the field "Return articles published in." You could also just use the `source:` tag before that phrase in the search bar: `source:"journal of marriage and family"`. To find a source from a particular website, use the `site:` tag instead (e.g., `site:nber.org`).
4. **Search for particular authors.** As you scroll through the list of articles in your search results, you should begin to notice that certain authors keep appearing. If you find an author that has written multiple articles on your topic, consider searching the author field for that particular author (in Google Scholar, use the `author:` tag followed by the author's name). You can also search the web for that author's curriculum vitae (CV), which will list their publications. Many academics maintain personal websites or put their CV on their university department's website. Just type in their name and "CV" into a search engine.
5. **Talk to a librarian.** They are professional knowledge-gatherers, and if you are at a university or college, a librarian is often assigned to your department or school. They routinely help students and faculty find the right sources for literature reviews.

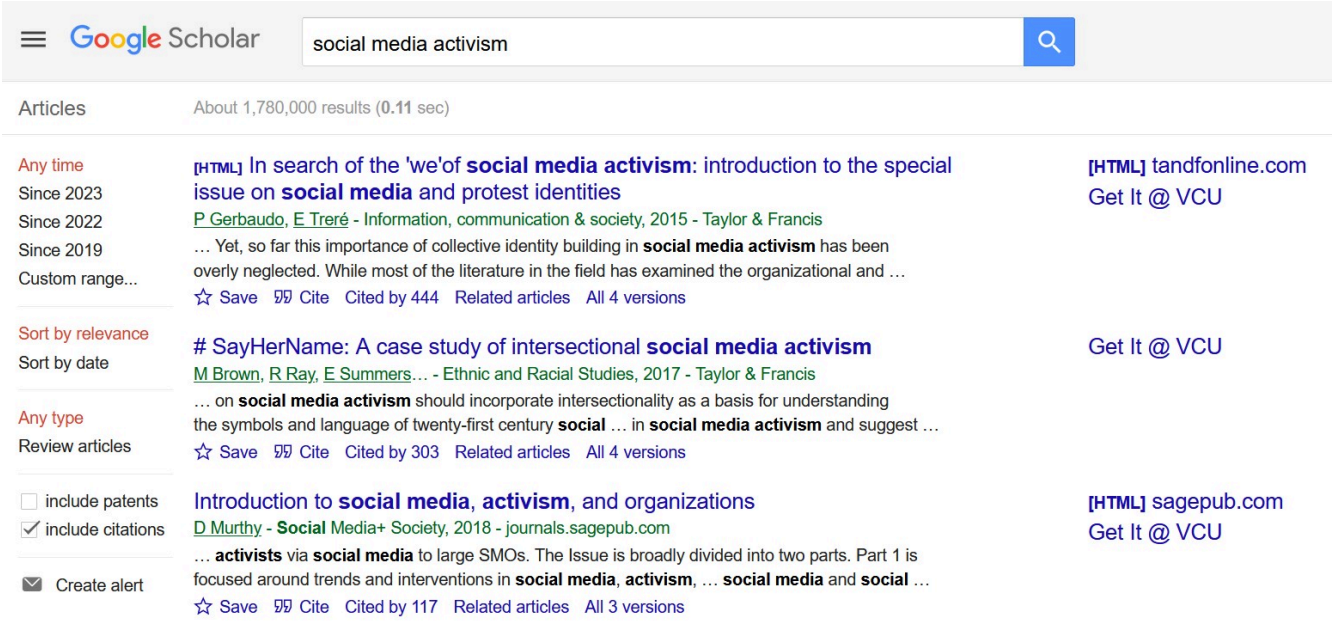


Figure 5.1. Google Scholar. When using Google Scholar for literature searches, you can easily narrow the search to cover certain years or certain types of articles (like review articles). In a desktop browser, these links are located by default on the left of the screen whenever you are looking at search results, and you can also click on the hamburger menu to select Advanced Search options. Note that you can connect your university library's login to Google Scholar using the Library Links option in its settings, which will create links to the right of search results that allow you to access those sources via your library's catalog with one click.

Earlier, we pointed out that it is a good idea to start your literature search by consulting a review article. Review articles are regularly published in journals like the *Annual Review of Sociology* and *Sociology Compass*, so you can search those journals in Google Scholar (e.g., `source:"annual review of sociology"`) or other databases to find one on your topic. On its results page, Google Scholar also has an option to narrow your search just to "Review articles" (see the link on the left-hand side of the screen in **Figure 5.1**). And if that option isn't available in the database you're using, you can include keywords in your query like "review article," "review of the literature," "state of the art," and "research agenda" to cover the same ground.

Likewise, you can use "systematic review," "meta-analysis," or "meta-synthesis" in your search queries to find those types of articles. Note that the Campbell Collaboration and Cochrane—two international research networks—produce systematic reviews on a variety of topics. Cochrane synthesizes medical studies to help physicians make evidence-based decisions in health care; the Campbell Collaboration publishes an open-access journal, *Campbell Systematic Reviews*, which evaluates social interventions in a variety of research areas: business and management, climate solutions, crime and justice, disability, education, international development, knowledge translation and implementation, methods, and social welfare.

Search Operators

Many databases (including Google Scholar) allow you to use *Boolean operators* such as `AND`, `OR`, and `NOT` to

connect words or phrases in your search query. This will easily narrow or broaden your search. An `AND` connecting two terms means both terms must be present, reducing the number of results you will get from your search query. For example, searching for `cocaine AND parenting` would return only those articles that mention both cocaine and parenting. (In some search engines, just listing both terms in your query—`cocaine parenting`—does the same thing as including `AND`.)

The `NOT` operator also reduces the number of results you get from your query. For example, perhaps you wanted to exclude issues related to pregnancy. Searching for `cocaine AND parenting NOT pregnancy` would exclude articles that mentioned pregnancy from your results. For the same effect, you can use a hyphen (–) right in front of the term you wish to exclude: `cocaine parenting -pregnancy`.

Using the `OR` operator can cover multiple phrasings, increasing the number of results you get from your query. This is helpful because researchers may use different words to mean the same thing—for example, using the terms "substance abuse" and "addiction." In this case, searching for `"substance abuse" OR addiction` would return all articles that mention one or the other search terms—in addition to those that mention both.

Step 3: Skimming Articles

Once you've found one or more articles that seem to be relevant to your topic, you need to see how they might fit in your literature review. Rather than read every article, read their abstracts and see which ones you really need to read. Abstracts appear in most reports of scholarly research and provide you with an easily digestible review of a study's major findings and the theoretical framework the author is using to position their findings. They typically contain just a few hundred words, so reading them will quickly familiarize you with a study.

A well-written abstract will give you both a framework for understanding the rest of the article and its punchline—what the author found and whether the article is relevant to your area of inquiry. For this reason, we suggest *skimming* abstracts as part of the literature search process. Use the keywords you find in them (sometimes also listed separately below the abstract) to continue your search for more relevant studies. If the study seems relevant to your paper, it's probably worth reading more. If it's not, then you have only spent a minute or so reading the abstract. Another way to get a quick snapshot of the article is to scan the headings, tables, and figures throughout the paper.

If you conclude an article is relevant, make sure to download the PDF or another version of the full text to your computer so you can read it later. You can keep the PDFs in folders on a cloud storage drive, arranged by topic. Better yet, you can use a reference manager like Zotero, which can help keep your sources and notes organized both on the cloud and on your computer. Many of these programs allow you to mark up those stored files with highlights and comments. Regardless of how you organize and annotate your sources, you do want to take notes and think about how each source might be of use in your study, as we describe in detail later.

Step 4: Examining References

As you begin to read your articles or books, you'll notice that the authors cite additional sources that are likely relevant to your topic area. You should definitely select sources for your literature review by looking through these reference lists. Follow the trail of citations from one source to another. See which papers are consistently cited across the sources you're reading—those are key studies that probably belong in your literature review, too.

This process of searching through reference lists is called **archival searching**. Unfortunately, this process allows you only to see relevant sources from before the citing article's publication date. For instance, the reference section of an

article from 2020 will include only references from before 2020. Some databases, however, allow you to do a forward-looking archival search, too. For instance, look up an article on Google Scholar and click the "Cited by" link at the bottom of its listing (refer to **Figure 5.1**). This is a list of all the articles that cite the article you just read. Google Scholar even allows you to search within the "Cited by" articles to narrow down your search further to articles that are more recent or relevant to your topic area. One good strategy is to find the seminal articles (the most highly cited ones) on your topic through your archival searching, and then see which newly published articles cite them. That way, you are capturing both the foundational research on your research question and the state-of-the-art studies on it. (For more on archival searching, check out this tutorial.)

Key Takeaways

1. When identifying and reading relevant literature, be broad in your search for articles, but be narrower in your reading of articles.
2. Conducting a literature search involves the skillful use of keywords to find relevant articles.
3. It is important to narrow down the number of articles in your search results to only those articles that are most relevant to your inquiry.

5.4. How Do I Read and Evaluate Sources?

Learning Objectives

1. Critically evaluate the sources of the information you have found.
2. Use the information you have gained from your literature review to identify a research problem your own study can address.

By now, you've hopefully collected a number of academic sources relevant to your topic area. It's now time to evaluate the information you've found. Using inaccurate, disreputable, and carelessly researched information will undermine the perceived quality of your own work. You not only want to be sure of the credibility and quality of each of the sources you use, but you also want to determine whether it is an appropriate fit for your literature review. You should always have a good reason to mention a work; if it does not help you make the case for your research problem and the study that addresses it, then cut it.

When evaluating whether or not to include a source in your literature review, ask yourself a series of questions:

1. **Does the author have the credentials to write on the topic?** Search the author's name in a general search engine like Google. What are the researcher's academic credentials? What else has this author written? Search by the author's name in the databases you are using (in Google Scholar, prominent authors will have their own pages) and see how much they have published on any given subject. If you are writing a sociological literature review, you typically want to prioritize authors who are sociologists, because then other sociologists who read your work will know whom you are talking about and will be able to clearly connect the studies mentioned to one another.
2. **Who published the source?** You will want to make sure that any journal articles you cite have been published by a reputable journal with high standards of peer review. Avoid predatory journals that publish anything, or journals that other scholars have not heard of—which typically means that well-respected scholars are not providing peer reviews for these journals. Likewise, you may want to prioritize books published by reputable publishers. Remember that university presses conduct peer reviews of the books they publish, whereas trade presses typically do not. For gray literature produced by think tanks, government agencies, or other organizations, look them up online to learn more about what potential biases they might have, including from their funding sources. Again, you generally do not want to include nonacademic sources in a literature review, but if you choose to do so, definitely avoid articles from obscure publications. A site with numerous typos, outdated design, dubious advertising, and other evidence of unprofessionalism is unlikely to provide trustworthy content.
3. **Is the source reliable and objective?** What point of view does the author represent? Are they upfront about that point of view? Is the article trying to argue a position, and if so, is it doing that in a way that adequately considers opposing sides? Is the article in a publication with a particular editorial position? Does it fairly present the findings or views of other authors with different perspectives on your topic? Do the researchers have any discernible conflicts of interest? On this last point, funders usually require that recipients acknowledge them in academic publications. But publications in the popular press routinely leave out funding sources, and the researcher may fail to disclose a personal stake they have in the research they conduct. In these cases, do a quick search on the web to see if you can learn more about a researcher's funding, affiliations, and other potential biases. You will probably want to put less weight in your literature review on research that appears motivated to support a particular

agenda. At the same time, the fact that a study received funding from a particular organization, even a political organization, doesn't necessarily discount its findings. And researchers may not be able to disclose certain details of their study even if they wanted to: as we will discuss in **Chapter 8: Ethics**, they are ethically bound to protect the identities of participants unless those people explicitly say otherwise. Nevertheless, getting as much context about a study as you can allows you to better assess what you can legitimately take away from a given set of findings.

4. **Is the information up to date?** Prioritize sources that are more current—published, say, within the past five or 10 years. It's fine to cite earlier work, especially since the more recent work builds upon that past work, but you want to make sure that the literature you discuss also covers what we currently know about the topic—not just what we used to know.
5. **How accurate is the source's information, and how strong is its evidence**? Check the facts in the article. Can the statistics be verified through other sources? Does this information seem to fit with what you have read in other sources? When sources mention existing research, do they accurately summarize the results of studies that you can read for yourself? For a research study, what research methods were used? How credible are its findings, given its research design? (On this last point, see the sidebar **A Hierarchy of Evidence?**)
6. **How important is this source in the literature?** As noted earlier, Google Scholar and other databases list the number of citations a particular source has within the scholarly literature—a rough measure of that study's impact within the research community. Your literature review might want to cover any papers that score high on these or similar indicators, since they are likely to be seminal within your field.
7. **Is the source relevant to your topic?** This is perhaps the most important question to ask. Some sources may be included in your literature review even if you have problems with their methods or findings—you can always explicitly discuss those problems, after all—but you don't want to waste your time reading and summarizing irrelevant literature. Don't try to read everything you dredge up during your searches, and don't include a source in your literature review unless you have a good reason to do so. To suss out a source's relevance, consider these questions: How does the article fit into the scope of the literature on this topic? Is it a general work that provides an overview, or is it focused on only one aspect of your topic? Does the information in it help you answer your research question, or is it difficult to make some sort of connection? Does the source present one or more points of view in a debate, so that you can show that you have addressed all sides of the argument in your paper?

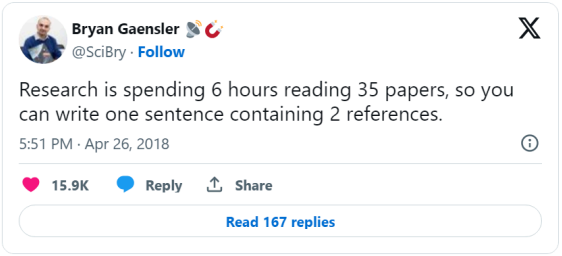
A Hierarchy of Evidence?

Sometimes researchers talk of a "hierarchy of evidence"—a view that certain types of studies produce findings that are more credible. Generally speaking, social scientists have greater confidence in findings that have been replicated multiple times. This is why meta-analyses can be so influential, as they use systematic procedures to synthesize the results of many studies on a topic. For single studies, experiments are considered the gold standard for deductive studies that seek to assess causality—that is, whether changes in the independent variable of interest *really* cause changes in the dependent variable of interest. As we will describe further in **Chapter 12: Experiments**, they have experimental and control groups, which allow researchers to isolate the effect of the independent variable on the dependent variable. Longitudinal studies follow a group of people to identify how variables of interest change over time, and therefore also allow researchers to see more directly the effect of an independent variable on a dependent variable. As a result of these features of their research design, the findings of experiments and longitudinal studies are often seen as more trustworthy than cross-sectional observational studies, which limit researchers in making observations at one point in time and which complicate any arguments about causal relationships.

This last point comes with some pretty strong caveats, as all studies have weaknesses and even cross-sectional studies can be useful. As we pointed out in **Chapter 4: Research Questions**, qualitative studies with small samples also play an important role in generating novel theories and refining them based on instructive outliers. In a field like sociology that employs such diverse approaches to research, it is usually a good idea to bring many different types of studies into your literature review. And regardless of the methods you are proposing for your study, your literature review should definitely cover any research that uses similar methods to investigate your research question. If you are conducting a qualitative study, for instance, you should include ample qualitative studies in your literature review so you can understand how others have studied the topic before you.

Breaking Down an Empirical Paper

While you should get into the habit of familiarizing yourself with each part of the papers you wish to cite, there are techniques for reading academic sources that can make them a little easier to digest quickly. Read the sidebar **How to Read Social Science Writing Strategically** for some general tips. During your initial searching of the literature, you should just be reading abstracts and perhaps skimming for some additional key details. Once you have identified an empirical paper that you'd like to read more carefully, skim through its introduction, literature review, discussion, and conclusion sections. (As you know, theoretical articles and books have no set structure, but reading the introduction and conclusion should cover much the same ground.) The paper's literature review will tell you what previous empirical and theoretical work the author considers to be important within this research area. The discussion and conclusion sections will help you understand what the author views as their study's major findings and limitations and how the author perceives their work to relate to other research.



Reviewing the literature is a labor-intensive task, but it is crucial to ensure that your work is actually scientific—that it contributes to existing knowledge. Thankfully, there are strategies to make your literature searching and reading much more effective, as we describe throughout this chapter. X

Ask yourself the following questions about each section of an article—from the abstract to the acknowledgments—as you read through it:

1. **Title:** What is the focus of this study? The key concepts in the study's research question (or the research question itself) will usually be included in its title. (Sometimes, the title is a sexy title meant to grab people's attention, while the subtitle contains the actual concepts of interest.)
2. **Author(s):** Who conducted the study, and what institutions are they affiliated with? What biases might have affected this research because of who the authors are or which organization they work for?
3. **Abstract:** What are the key findings? How were those findings reached? What theoretical framework does the researcher employ? Note that sometimes a journal article or book chapter will have keywords listed below the abstract that are used to help with database searching. These keywords will likely include the most important concepts in the study.
4. **Introduction:** How does the author frame their research focus? What other possible ways of framing the problem exist? Why might the author have chosen this particular way of framing the problem?

5. **Literature review:** How selective does the researcher appear to have been in identifying relevant literature to discuss? Does the review of literature appear appropriately extensive? Does the researcher provide a critical review? Carefully go through the paper's literature review as well as its reference list to see which other scholars the study's authors are in conversation with—whom they are relying on to support their arguments, and whose arguments they are countering, contradicting, or expanding upon.
6. **Methods (sample, case selection):** Where was the data collected? Did the researcher collect their own data or use someone else's data? What population is the study trying to make claims about, and does the sample represent that population well? What are the sample's major strengths and major weaknesses?
7. **Methods (data collection):** How was the data collected? What do you know about the relative strengths and weaknesses of the method employed? What other methods of data collection might have been employed, and why was this particular method employed? What do you know about the data collection strategy and instruments (e.g., questions asked, locations observed)? What don't you know about the data collection strategy and instruments?
8. **Methods (data analysis):** How was the data analyzed? Is there enough information provided for you to feel confident that the proper analytic procedures were employed?
9. **Findings or results:** What are the study's major findings? Are the findings linked to previously described research questions, objectives, hypotheses, and literature? Are sufficient amounts of data (e.g., quotes and observations in qualitative work, statistics in quantitative work) provided to support the conclusions drawn? Are the tables readable?
10. **Discussion, implications, or conclusion:** Does the author generalize the study's findings to some population beyond the sample used in the study? How are these claims presented? Are the claims made supported by data provided in the results section (e.g., supporting quotes, statistical significance)? Have any limitations of the study been fully disclosed and adequately addressed? Are the study's theoretical and policy implications sufficiently explored?
11. **Acknowledgments and declaration of conflicts of interest:** Who provided feedback on this research? Who provided support in the form of funding or other resources? What potential conflicts of interest did the researcher disclose? The declaration of conflicts of interest does not appear in every journal article, but if something is listed there, you should definitely consider the potential for bias in the research. While the acknowledgments may seem like a personal note, they can also be useful to help situate the work. Specifically, they help identify the study's major stakeholders. For instance, scholars will thank the foundations, government agencies, and other organizations that funded their research, which may say something about the scholarly or political perspective the authors are taking. Scholars often show their work to mentors or peers before submitting their work for publication, and they will thank these individuals in their acknowledgments. Whom they thank may give you clues about where they fall in the debates on a particular scholarly topic, especially when the authors are more junior scholars whose perspective may have been shaped by an influential mentor.

(*Note:* For a more detailed discussion of each part of an empirical journal article, refer to **Section A.3. Writing Up Your Work for Other Scholars**.)



As you read through sources, write down key details about each study, such as the research question it seeks to answer, the research problem (or gap in the literature) it addresses, its methodological approach and sample characteristics (including any limitations they present), and any findings and theoretical implications of the study that are particularly relevant to your own research. *Ono Kosuki, via Pexels*

As you read through the literature, take detailed notes and use a consistent format to record key information about each source. You will want to get through each paper quickly but also get everything you need. One strategy is to break down articles into their constituent parts, taking notes on the following details:

1. **Research question:** Read through the abstract, introduction, and literature review to identify the study's research question or questions. Note that the question may be phrased in multiple ways. In fact, you may wish to rephrase it for your own sake using the templates we described in **Chapter 4: Research Questions**, which may clarify for you the variables and relationships that are central to the study. For a qualitative study, the research question could take the form: "How does the independent variable affect the dependent variable?" For a quantitative study, it could take the form: "What effect does the independent variable have on the dependent variable?" In any case, make sure to identify all the relevant variables as well as how they are measured—not just those stated explicitly in the research question, but also (if applicable) moderating variables, control variables, and causal mechanisms.
2. **Research problem:** Look to the abstract, introduction, and literature review to find the study's research problem. If the paper was clearly written, it will talk about the specific gaps in the literature that the study intended to address. You may come across phrasings that signal a research problem: "Past research assumed," "Scholars have not yet looked at," "What has been less studied is," and so on. (See the University of Manchester's useful Academic Phrasebank for other telltale language that scholars use to highlight the "inadequacies of previous studies," the "paucity or lack of previous research," or a "knowledge gap in the field of study"—that is, how they flag potential research problems.) Watch for these turns of phrase in the articles you read, which will help you spot the research problem that each study is seeking to resolve.
3. **Hypotheses:** These hypotheses should be the expected answers to the research questions, derived from the study's literature review. They might be listed formally or described in paragraph form, and sometimes (particularly for qualitative work) they are implied rather than explicitly stated.
4. **Methods:** The paper's methods section (which may go under different names and may encompass multiple sections) will tell you about the type of data collected (units of analysis), the sample size, the sample selection procedures, data collection and data analysis methods used, the measures used (i.e., how the study's key concepts were operationalized), the control variables used, and so on. If the study is particularly important to your own study—perhaps because you are improving upon its methods, or studying the same question in a different context—you may want to write extensive notes about the methods it used; otherwise, you can quickly summarize the highlights.
5. **Findings:** Based on summaries throughout the article (including the abstract, introduction, and conclusion), you should be able to boil down the study's results to several key takeaways. These should be answers to (1), the research question, and may or may not coincide with (3), the hypotheses (you should note whether the findings confirm or disconfirm those hypotheses). In addition to summarizing the study's overarching arguments, it's a good idea to include specific examples—of key statistics from a quantitative analysis, or representative quotes or vignettes from qualitative analysis—to flesh out those points in your notes. Quantitative studies may concisely capture some of these takeaways in their tables and charts.
6. **Theoretical implications:** The abstract, introduction, discussion, and conclusion should make a case for how the study addressed (2), the research problem. Circle back to the research problem and see how exactly the study built on, contradicted, or otherwise engaged with past studies. Also, see if the authors discuss directions for future research (e.g., "more research is needed in this area"). They usually do this at the end of their paper, sometimes in a separate section called "Future Research." Record any promising ideas, which may flag a research problem that

your own study can address.

7. **Limitations:** The authors should have discussed the limitations of their work in the paper's conclusion or methods section or in a separate limitations section. If possible, you should go beyond what limitations the authors divulged and see if you can identify any strengths and weaknesses in the methodological or theoretical approaches they pursued for their study. Are there any **scope conditions** for the study—that is, contexts where its findings are likely to hold up or not hold up? Were there any shortcomings in their data collection or analysis? Every social scientific study has weaknesses, and you should be able to identify one or more just by the methods they chose—all of which have trade-offs, as we have discussed.
8. **Other important details:** For articles that are highly relevant to your study, you may want to record other specifics not captured in these categories—for instance, the definitions of key terms, or any statistics or trends mentioned that might provide helpful context.
9. **Relevance to your research question:** Finally, note how the source relates to other studies, and how it is relevant to your study. (The two considerations are intertwined, as your study may want to address disagreements between this study and others, or build on that previous body of work.) You may want to rank each source as "high," "medium," or "low" relevance. For papers that you decide not to include in your literature review, you may want to note your reasoning for exclusion, such as "small sample size," "local case study," or "lacks evidence to support conclusions."

Depending on your personal preference, you might want to store your notes on each source in a grid, table, or an outline. This breakdown (sometimes called a **literature review matrix** or *summary table*) will help you compare and contrast sources, see the relationships among them, and identify common themes as well as areas where you may need to do more searching. You can organize this information within a document or a spreadsheet—perhaps ordering it by the author's last name, with each source on a separate page in a document or a separate row in a spreadsheet. The spreadsheet version of this literature review matrix is depicted in **Figure 5.2**. Whichever method you choose, this type of organization will help you to both understand the information you find and structure the writing of your review. Remember that the goal is always to relate whatever you read specifically *to your research question.*

	A	B	C	D	E	F	G	H	I
1	**Citation**	**Research question**	**Research problem**	**Hypotheses**	**Methods**	**Findings**	**Implications**	**Other details**	**Relevance**
2									
3									
4									
5									
6									
7									
8									
9									
10									
11									
12									
13									
14									
15									
16									
17									

Figure 5.2. Literature Review Matrix. You can use a spreadsheet or table to capture the essential details about each of the sources you read for your literature review. This breakdown of each article will allow you to see connections between studies and focus your attention on their relevance to your driving research question. *Kelsey Cheshire, via VCU Libraries*

As you read through your expanding matrix of sources, look for common themes, as they may provide the structure for your literature review. And, remember, research is an iterative process. You may find yourself going back and searching reference lists and databases for more sources as you process the articles you've already collected.

How to Read Social Science Writing Strategically

Let's face it: social science writing is not Shakespeare. Especially as you become more familiar with a particular topic, you will not want to read studies on that topic word for word. You won't remember many of the details afterward, anyways, and you might lose yourself amid all the details and find yourself, hours later, drooling on your desk, with no idea what the text is saying. Instead, we recommend you read strategically. Let's go over one efficient approach to reading social science writing. Note, however, that you should do whatever works best for you personally.

First, figure out exactly what you would like to get out of the reading. Is the point to critique it for a class discussion? Is it to obtain concrete examples to back up a point in a paper you're writing? Is it to get a sense of what research has been done on a topic? Depending on your goals, you'll want to focus on specific parts of the reading. See the other sections of the current chapter and **Section A.3. Writing Up Your Work for Other Scholars** for a breakdown of the parts of a journal article and what they contain. Also, these guidelines (Laubepin 2013) have a handy chart telling you where to look in a paper given what you want to get out of it.

With the goal of your reading in mind, skim the article or chapter (5 to 10 minutes) to get a general sense of the material. Don't read every word. Read the introduction and conclusion, skim the section headings and the first sentence of each paragraph, look at any charts or tables, and so on. Now that you have a sense of the structure of the reading and its overarching points, you can home in on the portions of the text that are most useful to you, reading them more closely.

Let's say you don't have a very focused goal in mind for your reading. You just want to digest the information, perhaps to help you with a literature review. In that case, it's a good idea after your initial skim to write a sentence or two that summarizes the research question and research problem the article or chapter is trying to address and its main arguments or findings.

Then read more carefully, fleshing out in your notes what you think are the most important terms, arguments, examples, and trends, as well as any criticisms, questions, or thoughts that the reading provokes in you. Keep in mind the research question and research problem you identified earlier and connect anything you read and write down to those points. Feel free to skip around or even read the various sections of the article or chapter out of order. Pay special attention to the introduction and conclusion and any tables or charts (a table or chart will often summarize the entire argument of a quantitative study). Write down *selected* examples, anecdotes, and statistics to illustrate the main points you have identified in your short summary, and make sure that you know the definitions of all the key concepts in the reading. That said, your goal is to understand the article or chapter's most useful and interesting points and their connections to the rest of the piece and to other readings. You are not reading passively with the aim of remembering all the details.

After this second reading of the material, refine your summary and (especially if you are reading the text for a literature review) write another sentence or two describing the work's relevance to your research question or topic of interest and how it connects to other relevant scholarship. In the end, you will have a one-paragraph summary of the study and a one-paragraph description of its relevance to your own research interests, in addition to your unstructured notes from your closer reading. If you are pressed for time, prioritize these two

paragraphs over the detailed notes. (By the way, some educators recommend the Cornell note-taking method, which will reinforce the main points you actually want to remember.)

How do you tackle an entire book? First of all, note that many academic journals publish book reviews of monographs, which you can find easily using the online search methods described in this chapter. You may want to read one or more book reviews before cracking open the book itself. Much like the initial skim of a paper or chapter, reading a book review can help you figure out the structure of a book and its key arguments, allowing you to do a more focused reading of the portions most relevant to you. You can read this article (Lurie 2014) for more tips, but please be advised that, contrary to the tantalizing "read a book in two hours" headline, your mileage may vary.

Note that this sort of actively engaged reading will always be a balancing act. On the one hand, you want to get through a particular reading in a reasonable amount of time, and with a good understanding of its *overall* argument, rather than getting bogged down in *minor* details (for instance, specific statistics—unless, again, that's your primary purpose for doing the reading!). On the other hand, in academic reading you also need to remember enough detail—especially tangible examples—to flesh out those broader arguments in concrete and memorable ways. Strategic reading is a skill, like any other, that takes time to master.

Identifying a Research Problem

By now, we have hammered into your head that your literature review must identify a research problem and must make a strong case that your study will fill that gap in the literature. But how exactly does it fill the gap? In a nutshell, a study addresses a research problem by *agreeing* or *disagreeing* with past research.

Wayne Booth and his collaborators (2016) describe these two stances as *creative agreement* and *creative disagreement*. In the first case, a study builds on existing work—for instance, by providing new evidence to confirm its findings, extending its theories to new contexts, or fleshing out particular aspects of its explanations. In the second case, a study contradicts past research, such as by finding alternative explanations for a phenomenon or challenging the evidence or logic of previous studies. This agreement or disagreement generally takes one of three forms. First, the new study can improve theoretically on previous scholarship, bringing new or refined theories to bear on the phenomenon of interest. It can improve empirically on past work, introducing novel or superior data. And it can improve on the literature methodologically, utilizing better methods of data collection and analysis.

There are a wide variety of strategies for creatively agreeing or disagreeing with other scholars, and you can look at Booth and his collaborators' textbook (2016) for a fuller accounting. We want to flag two relatively straightforward intellectual moves, however, that novice researchers can follow to identify a research problem and justify their proposed study. First, you can creatively agree with an existing theory by extending it to new contexts. For instance, previous research found that ride-sharing apps operate in an exploitative fashion in rich countries; does the same dynamic hold up in poor countries, where different institutions of government regulation and worker organizing are operating? Second, you can creatively disagree with previous work by applying a different theoretical perspective. Psychologists have looked at unemployment in terms of the ways it challenges a person's identity; can we look at this identity loss from a sociological perspective, in terms of how the breaking of ties to organizations and social networks weakens those identities?

The articles you read as part of your literature review will help you uncover one or more research problems, though you will also have to draw on your own creativity and critical thinking as a sociologist. Please be aware that the research problem that your study ultimately settles on may not be the one you started out with. Through the process of data collection and analysis, you may realize that your findings speak to a more important or pertinent research problem

than what you initially identified in your literature review. That's a good thing. Skilled researchers adapt to the facts on the ground, and they inevitably come up with more innovative and complex ideas as they dive more deeply into the data.

Key Takeaways

1. In your literature review, prioritize high-quality sources and discuss concerns with problematic sources, based on potential author biases and study limitations.
2. When reading a source, make notes about its research question, research problem, hypotheses, methods, findings, implications, and relevance to your study. You can organize these details in a literature review matrix to help you synthesize the sources you read.
3. As you read more articles, identify a research problem in the literature based on how your study might creatively agree or disagree with past work.

5.5. How Do I Write a Literature Review?



Think carefully about the structure of your literature review. Divide your material into themes, with each thematic section

grouping together "like" bits of information—for instance, similar sets of findings from the studies you are synthesizing. Think up a title for each section that makes a particular point or argument (i.e., a title that contains both nouns and verbs) rather than just identifying a broad idea. You don't actually have to use the titles as headers for your sections, but they will help focus your attention on the distinct takeaways of the different parts of your literature review. *Mariam Antadze, via Pexels*

Learning Objectives

1. Describe the material that goes into the introduction and main body of your literature review.
2. Learn how to organize the arguments in your literature review in sections, according to a funnel structure, and with adequate signposting.

At what point do you stop reading and start writing your literature review? That's a judgment call you have to make on your own. Although you can fill in material later, there is always the danger that you will start writing without really knowing what you are talking about or what you want to say. For example, if you don't have a solid definition of your key concepts or a sense of how the literature has developed over time, it will be difficult to make coherent scholarly claims about your topic in your literature review.

At the same time, we would encourage you to start the process of writing as soon as you can, and to be comfortable with having that process stop and start as you move back and forth between writing and reading more. Writing is the best way of thinking through an intellectual problem, and as you write, you will begin to recognize what bits of information are missing in your take on the existing literature and on the research problem you hope to address. For one thing, you will figure out how much detail you need about each of the studies you are reading through for your literature review. For studies that are not highly relevant to your own research question, you may just need to mention them in passing, and with that knowledge, you can avoid spending unnecessary time pouring through those texts.

We'll have more to say about writing a research proposal later, but at this point we want to mention that its first two sections—the introduction and literature review—will look much like they would for a stand-alone literature review. Furthermore, writing a research proposal—which is essentially a literature review plus a methods section—can be quite straightforward once you have a good sense of the literature and have identified a compelling research problem and research question.

The Introduction to Your Literature Review (or Research Proposal)

The introduction to your paper or proposal should define or identify the general topic, issue, or area of concern. Whether you're writing a full-length empirical article, a research proposal, or a stand-alone literature review, you should start off by establishing why the topic of the paper is a social problem worth studying. This pulls the reader into the paper: who would want to read about something unimportant? This "so what?" introduction to the paper generally answers the following questions, although this list is far from exhaustive:

1. Why is this an important problem to study?
2. How many people are affected by the problem?

3. How does this problem affect other social issues or target populations relevant to sociology?
4. Why is your target population an important one to study?

A strong introduction should be filled with facts, theory, and arguments based on the literature you've found. Do not start with your personal reflections or experiences relating to the topic. In fact, you don't want to mention yourself at all in the literature review, although it is fine to use first-person voice when describing what you plan to do (for a research proposal) or what you already did (for an empirical paper). And don't go on for too long; spend no more than a few paragraphs on the paper's introduction. In these paragraphs, you might present historical background, mention the results of a particularly important study (especially one central to your research question), and provide definitions of important terms.

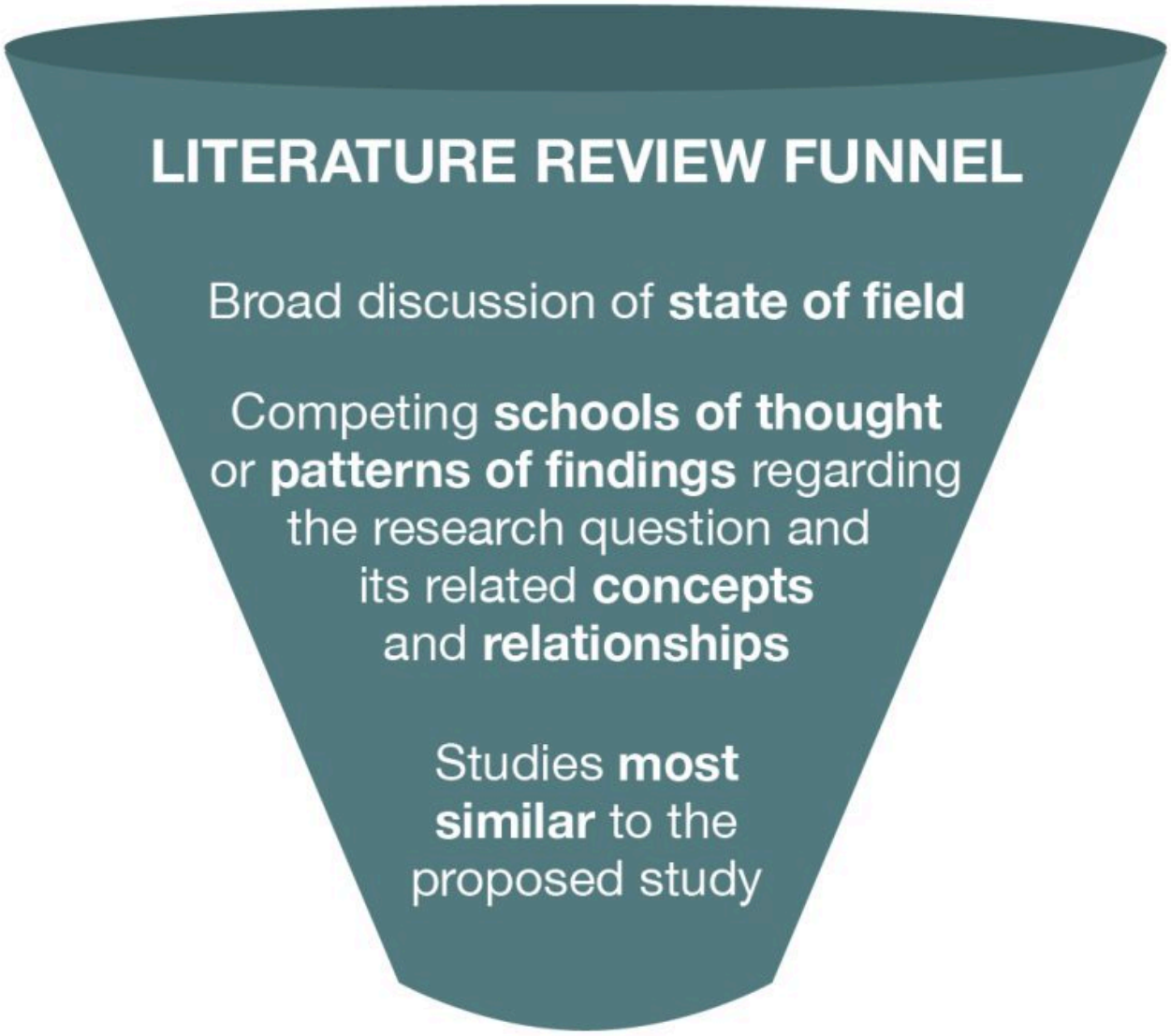
At the end of the introduction, you might want to include a few sentences that walk the reader through the rest of the literature review. In a stand-alone literature review, you should highlight your main arguments from the body of the literature review—for instance, what research problems currently exist in the literature, and how scholars should go about resolving them (i.e., the research agenda you're putting forward). In a proposed or completed study, the introduction should end with a statement (or restatement) of the research problem you seek to address, the research question that motivates your study, the methods used to answer that question, and the potential theoretical implications of the study's findings.

The Thematic Sections of Your Literature Review

After the introduction comes the actual literature review. Generally speaking, do not organize your literature review by source—that is, one paragraph for source A, one paragraph for source B, and so on. This is the approach of an annotated bibliography, not a literature review. You might be adequately *summarizing* the literature you've found, but you would not be *synthesizing* it. You would not be telling your reader how to put all those findings together, what points of agreement or contention exist in the literature, or how each study builds on the work of others.

As we've emphasized, your literature review is an argument for your research problem and the research question and study that address it. Like any effective argument, your literature review must have some kind of structure. For example, it might begin by describing a phenomenon in a general way along with several studies that provide some detail, then describing two or more competing theories of the phenomenon, and finally presenting a hypothesis to test one or more of the theories. Or, it might describe one phenomenon, then describe another phenomenon that seems inconsistent with the first one, then propose a theory that resolves the inconsistency, and finally present a hypothesis to test that theory. In applied research, it might describe a phenomenon or theory, then describe how that phenomenon or theory applies to some important real-world situation, and finally suggest a way to test whether it does, in fact, apply to that situation.

When writing up your literature review, it is extremely important to start with an outline of the main points that you want to make, organized in the order that you want to make them. Use your notes on the reading you did to help you identify the important topics you need to cover, and—more important—what you personally have to say about those topics. The basic structure of your argument then should be apparent from the outline. Unfortunately, no formula exists that will work for everyone. We can, however, provide some general pointers on structuring your literature review.



Conclude by stating or restating the **research problem** (the puzzle or gap in the literature), the **research question**, and/or any **hypotheses**

Figure 5.3. Organizing Your Literature Review as a Funnel. A good way of staging your literature review is by following a funnel structure. At the top, you broadly sketch the literature on your topic. Here, previous studies are often cited one after another, with little, if any, discussion of them individually. As the literature review proceeds, you bring up studies increasingly similar to your own, spending increasing amounts of space describing each one. Finally, you conclude the literature review by making the case for your research problem—that is, what those similar studies near the end have missed. When you have multiple sections in your literature review, the whole review may take the form of a funnel, or you may instead choose to structure each of your sections in that way (perhaps identifying multiple research problems across those sections).

First, you will find that many literature reviews have a "funnel" structure, with a broad discussion of the past literature at the top, and then an in-depth discussion of specific articles at the bottom (White 2005). Articles that closely relate to the authors' own study will be positioned closer to the end. This is a logical structure that naturally concludes the literature review with a statement of the research problem and any hypotheses.

Note that organizing your arguments in this way will affect how much you say about each source. In the intellectual performance that is a literature review, not all cited scholars have equal time on stage: you may devote an entire paragraph or more to discussing an author's work, or you may mention it in a single sentence, or in a long list of references that back up a broader point. How long you spend in discussing a particular study depends on how critical it is to making the case for your research problem and the study that addresses it. For instance, when your own study will try to extend or challenge another study's findings, you should discuss that previously published article in great detail—describing its methods and findings, flagging its shortcomings and limitations, and pointing out any questions it raises. In this case, you probably need to go well beyond the article's abstract in describing the study's research design, results, and implications. But if your goal is to acknowledge a section of the related literature that doesn't bear so heavily on your study, then you should just mention those studies in passing. (Per the funnel structure, the former study will go near the bottom of your literature review, since it is closely related to your own; the latter studies would go near the top.) Only describe aspects of a study that are relevant to your literature review. For instance, if you are studying the effect of social media on grassroots activism, you might briefly mention a study that looks at how social media affects political attitudes more generally. But you probably wouldn't need to say much, if anything, about that study's research design, given that your focus is grassroots activism rather than political attitudes. Don't insert extra details about a study just to take up space.

A second important piece of advice is to break up your literature review into sections. These sections (and possibly subsections) should be organized by themes—as opposed to chronologically or by author, for instance. You do not necessarily have to use explicit section headings—if you prefer, the sections can just be in the internal outline you're using when writing things up—but you do want to think of your literature review in these modular terms. That way, you can put "like" bits of material—sources that examine the relationships between the same variables, for instance, or come to the same conclusions—together in a single section. (Indeed, you can think of these sections as "buckets" in which you throw in all the studies that have to deal with one particular aspect or question relating to the phenomenon you're studying.) You can move the sections up or down depending on where they best fit, perhaps with the sections most directly relevant to your research question—that cover the exact same causal relationship, for instance—near the bottom. You may also want to begin each section with a new funnel argument, ending that section with another research problem that your survey of the literature in that particular area has identified.

The use of thematic sections places the focus of your literature review on patterns and trends in the literature, rather than on specific authors per se. That is what you want in your critical evaluation of the literature. Point out consistent findings and note exceptions and contradictions. Emphasize stronger studies over weaker ones, explaining how you evaluated them. For research that is closely related to your own study, take the time in your literature to describe their strengths and weaknesses—which may make the case for a research problem that your study can resolve empirically, theoretically, or methodologically.

Having a funnel structure to your literature review and breaking it down into thematic sections will help ensure that you

are making a coherent argument on behalf of your research problem and proposed study. Here are some additional tips for writing the body of your literature review:

1. **Use quotes sparingly.** Paraphrase when you need to relate specific details from a source, condensing and rewording the text (one trick is to read the passage in the original source and then look away from the text, forcing yourself to sum it up in your own words). Quotations for definitions are acceptable, but reserve quotes for when someone says something so well you couldn't possibly phrase it differently. Never use quotes for statistics.
2. **Stay objective and acknowledge disagreements in the literature:** In your literature review, you are constructing an argument for why your research question is interesting and worth addressing—not necessarily why your favorite answer to it is correct. Make sure that your discussion of past research is balanced. If you want to emphasize the generally accepted understanding of a phenomenon, then of course you should discuss various studies that have demonstrated it. However, if other studies have found contradictory findings, you should discuss them, too. Likewise, if you are proposing a new theory, then your literature review needs to cover not just the empirical research that supports your theory, but also any studies that failed to do so. It is acceptable to argue that the balance of the research supports the existence of a phenomenon or is consistent with a theory (and that is usually the best that researchers can hope for), but it is not acceptable to ignore contradictory evidence. In line with the objective stance you're taking in this section, do not approach the writing of the literature review as you would a personal essay or reflective journal entry—keep the tone formal, avoiding informal language like contractions, idioms, rhetorical questions, or first-person voice (for a research proposal, though, using "I" or "we" is fine when discussing your methods).
3. **Back up every claim you make.** Your arguments in the literature review need to be based on specific empirical facts or established theories. Use citations liberally in your literature review, backing up any claim that isn't patently obvious. Note any statements that lack evidence from the scientific literature on your topic; you may need to track down additional sources or at the very least qualify your statements (with words like "perhaps," "may," or "could") so that you are not overstating your case.
4. **But do make claims: you are the expert.** These cautions aside, however, remember that literature reviews are written from the perspective of an expert on the field. After an exhaustive review of the important studies on a topic, you should feel like you are able to make strong claims about what is true—so make them! Do not use terms like "I believe" or "I think"; simply state what you know of the literature.

In the last paragraph or so of the literature review section, you should briefly sum up the gaps in the literature you have identified and then state (or restate) your case on behalf of why your research question and study will solve that research problem. (If you wish, you can put this concluding argument in a stand-alone section called "Statement of the Problem," "Problem Statement," or "Research Problem.") If you are writing a stand-alone literature review that doesn't support your own proposed or completed study, this section can end with multiple recommendations for further research wherever you have identified gaps in the current literature. Be as specific as possible about the implications of your review of the literature—don't just say that more research is needed or that the topic is understudied, but rather point out how exactly the research problem could be addressed.

What happens if you run into writer's block? Type out whatever you want to say, regardless of how good it is. Even if you have a detailed outline to work from, the words are not going to fall into place perfectly the first time you start writing. You should consider turning off the editing and critiquing part of your brain for a little while and allow your thoughts to flow. Don't worry about putting the correct internal citation when you first write. Just get the information out. Only after you've reached a natural stopping point might you go back and edit your draft for grammar, formatting, organization, flow, and more. Divorcing the writing and editing process can go a long way to addressing writer's block—as can picking a topic about which you have something to say.

Signposting: Making Your Arguments Clearer



Using signposting in your academic writing will help your readers follow your argument and understand how the pieces of it—including your discussion of each study in your literature review—connect with one another. It can include using headings and subheadings, topic sentences, and transitions between paragraphs that make a critical point, such as a comparison or contrast you are drawing between examples. *Zoe Davidson, via Flickr*

To clarify the structure of your argument, you should engage in **signposting** throughout your literature review and research proposal (we also discuss signposting in the context of research presentations in **Appendix A: Presenting, Writing, and Publishing**). When you signpost, you highlight words or phrases to identify the organization and structure of your argument. The most basic form of signposting is using a topic sentence at the beginning of each paragraph. A topic sentence introduces the argument you plan to make in that paragraph. For example, you might start a paragraph stating, "There is strong disagreement in the literature as to whether educational attainment improves cognitive ability, or whether people with higher cognitive ability are more likely to go to school." Within that paragraph, your reader would likely assume you will present evidence for both arguments. The concluding sentence of your paragraph can return to the topic sentence, perhaps synthesizing the various arguments presented to make a specific conclusion.

Signposting also involves using headings and subheadings. Creating different sections—each with a heading that describes the theme or argument of that section—will help your reader understand the structure of your literature

review. Headings can also help if the reader gets lost and needs to reorient themselves within the document. It's like walking a small child around, telling them: "First we'll do this, then we'll do that, and when we're done, we'll know this!"

Another way to use signposting is to open each paragraph with a sentence that links the topic of the paragraph with the one before it. (Alternatively, one could end each paragraph with a sentence that links it with the next paragraph.) For example, imagine we wanted to link a paragraph about barriers to accessing health care with one about the relationship between the patient and physician. We could use a transition sentence like this: "Even if patients overcome these barriers to accessing care, the physician-patient relationship can create new barriers to positive health outcomes." A transition sentence like this builds a connection between two distinct topics. Transition sentences are also useful within paragraphs. They tell the reader how to consider one piece of information in light of previous information. Even simple transitions like "however," "similarly," and "nevertheless" demonstrate critical thinking and make your arguments clearer.

Many novice researchers have difficulty with incorporating transitions into their writing. Let's look at an example. Instead of beginning a sentence or paragraph by launching into a description of a study, such as "Williams (2004) found that . . . ," it is better to start by indicating something about why you are describing this particular study. Here are some simple examples:

- Another example of this phenomenon comes from the work of Williams (2004).
- Williams (2004) offers one explanation of this phenomenon.
- An alternative perspective has been provided by Williams (2004).

Key Takeaways

1. The introduction to a paper draws the reader into your topic by highlighting how important the topic is to sociology and overall society.
2. The literature review itself should be organized by themes, possibly with separate sections following a funnel structure that moves from a general discussion to an in-depth analysis of research closely related to your research question.
3. Signposting is an important component of academic writing that demonstrates critical thinking and helps guide your reader through your arguments.

5.6. How Do I Write a Research Proposal?

Learning Objectives

1. Identify which research designs may be useful for answering your specific research question.
2. Identify the aspects of feasibility that shape a researcher's ability to conduct research.

In the last chapter, we discussed what the stages in research design might look like:

1. Develop an initial research question—what is called a **working research question**, given that it is a work in progress (**Chapter 4: Research Questions**)
2. Decide on an overall research strategy—whether your analytical approach will be inductive or deductive, what the target population is, and so on. (**Chapter 4: Research Questions**)
3. Conduct a literature review to identify a gap in the existing body of research (a *research problem*) and refine your research question (**Chapter 5: Research Design**)
4. Decide on a method of data collection and analysis (**Chapter 5: Research Design** and the chapters devoted to each method)
5. Propose hypotheses based on your literature review (**Chapter 5: Research Design**)
6. Decide on your sampling strategy (**Chapter 6: Sampling**)
7. Define key concepts and measures—what is called **conceptualization** and **operationalization**, respectively (**Chapter 7: Measuring the Social World**)
8. Identify and address any ethical concerns about your proposed study (**Chapter 8: Ethics**).

As we noted, decisions about research design do not necessarily occur in sequential order. For example, you might have to think about potential ethical concerns before zeroing in on a specific research question or sample, since such considerations will limit whom or what you can study. Similarly, your method of data collection might hinge on what population of interest you want to generalize about—or vice versa. As you design your research project, you'll be constantly moving back and forth between steps on the list below. Sometimes researchers even jump ahead to data collection and analysis, and then circle back to refine their research question. Although not ideal, this approach sometimes makes practical sense. Maybe you're using an existing dataset and need to know what concepts are adequately operationalized within it before you decide on what to study. Or maybe you need to talk to some people before you can identify and read up on the underlying variables driving a particular relationship that your research question examines.

All that said, a good literature review sets you up nicely for choosing among possible methods of data collection and analysis. You know what sorts of research have already been done on your topic, and you have identified a compelling research problem. Now you need to decide on a methodological approach to answer your research question and address your research problem. We will learn more about the following methods in the remainder of this textbook:

1. Ethnographic observation
2. In-depth interviews and focus groups
3. Experiments

4. Surveys
5. Historical analysis
6. Content analysis
7. Social network analysis

So many different research designs exist in the social scientific literature that it would be impossible to include them all in this textbook. We cover the most popular methodological approaches in sociology, which we hope will give you a foundation for understanding more advanced designs as well. We encourage you to learn about other methods and more specialized study designs not reviewed in this textbook, which will expand your toolkit for conducting solid research.

In this section, we'll discuss how to write a research proposal, which is a plan for the design of your study. The first consideration is to decide on a general methodological approach.

Choosing a General Methodological Approach

The rest of this textbook is devoted to describing specific research methods, so we will have much more to say about each approach and its advantages and disadvantages. At this point, we want to make some general comments about choosing between qualitative and quantitative methods. As we have noted, the distinction between qualitative and quantitative research methods boils down to the *kind* of data we sociologists collect and analyze. Quantitative research employs numeric data, such as scores and metrics. Qualitative research relies mostly on nonnumeric data, such as quotations from interviews and details from visual observations. As a result, qualitative research is not well-suited to be analyzed with statistical procedures. Sometimes, qualitative data is tabulated quantitatively—based on how frequently certain codes come up, for example. Many qualitative researchers reject this coding approach, however, as a futile effort to seek consensus or objectivity in a research approach that is essentially subjective.

The strength of qualitative methods, as we've emphasized, is in the richness of its data—the fact that stories and meanings are not collapsed into numbers but captured in all their complexity and detail. This **thick description** creates a sound empirical foundation for theorizing about how causal mechanisms operate and how people understand their social realities. The most frequently used qualitative research method is **in-depth interviews** (face-to-face, phone, or online, conducted with individuals or focus groups). A second technique is **ethnographic observation**. Observational techniques include **bystander observation**, where the researcher is a neutral and passive external observer, and **participant observation**, where the researcher is an active participant in the phenomenon and their input or mere presence may influence the phenomenon being studied. When qualitative researchers talk about being "in the field," they're talking about using one or both of these approaches to be out in the real world and involved in the everyday lives of the people they are studying. A third qualitative technique is qualitative **content analysis** of documents (books, emails, annual reports, financial statements, news articles, websites, etc.) or media (videos, movies, songs, etc.), which does not have to occur in the field.

In sociology, quantitative research is usually conducted using **surveys**. Typically, a researcher creates a quantitative dataset by asking people to answer multiple-choice survey questions and then counting their responses. (Alternatively, the researcher analyzes an existing dataset based on someone else's survey.) Using scientific sampling procedures, sociologists can be confident that the subset of people they interviewed more or less represents the population they want to know about. That means the results of the survey—the percentage of people who had a certain political opinion, for instance—can be generalized to that larger population. Furthermore, because of the ways that quantitative measures simplify reality—boiling down wide-ranging opinions on a particular policy to several response categories, for instance—surveys allow people's responses to be easily compared across groups, time periods, and the like. They also allow for the strength of relationships between variables to be calculated: for instance, how much a person's gender affects how much money they make doing the same job, or how much being born in a certain part of the country affects a person's happiness and health. Quantitative content analysis and **social network analysis** apply similar approaches to

other forms of data (text and media for the former, and network data for the latter), tallying up phenomena in ways that can ideally say something meaningful about larger populations. (Our textbook does not currently cover social network analysis, but you can get a sense of what this methodological approach covers in the sidebar **Mapping the Social World**.)

While less common in sociology than in sister fields like psychology, **experiments** trade some generalizability for greater confidence about causality. By having control and experimental groups that are exactly the same except for the presence of a stimulus or treatment, experiments allow researchers to infer more directly that a change in one variable leads to a change in another. They can also quantify the size of that effect—in this case, with more certainty that it is causal.

For all these reasons, quantitative methods add more precision and universality to our understanding of a phenomenon than qualitative methods are capable of. At the same time, the quantification of data inevitably strips away nuance and complexity, possibly flattening meanings and oversimplifying relationships. To avoid the pitfalls of each approach, a **mixed-methods research design** uses qualitative and quantitative techniques jointly within a single study, with the hope that multiple methods can complement and compensate for one another.

So what sort of research design should you pursue for your study? Ideally, your research design is determined by your research question. If your research question involves, for example, testing a new policy intervention, you will likely want to use a deductive experimental design. If you want to know the lived experience of people in a public housing building, you probably want to use inductive qualitative methods like in-depth interviews and ethnographic observation. In other words, you should pursue a research design that gives you the best chance of arriving at a compelling answer to your research question. We'll talk more about which research methods are best at answering which research questions in later chapters.

As a researcher, you have to choose a research design that not only makes sense for answering your research question but also is feasible to complete with the skills and resources you have. For one thing, the design of your research study determines what you and your research participants will do. In an experiment, for example, the researcher will introduce a stimulus or treatment to participants and measure their responses. A content analysis may not have participants at all, and the researcher may simply be reviewing organizational materials or news media to understand cultures and attitudes. Your personal preferences and talents working with data and other people will naturally lead you to pursue certain methods of research.

All research projects also require resources to accomplish. Make sure your design is one you can carry out with the time, money, and assistance available to you. For instance, are you interested in better understanding the day-to-day experiences of maximum-security prisoners? This sounds fascinating, but unless you plan to commit a crime that lands you in a maximum-security prison (generally not a wise choice even for an ambitious researcher), gaining access to that facility would be difficult. Perhaps your interest is in the inner workings of toddler peer groups. If you're much older than four or five, however, it might be tough for you to access that sort of group. Your ideal research topic might require you to live on a chartered sailboat in the Bahamas for a few years, but unless you have unlimited funding, it will be difficult to make even that happen. While the topics about which research questions can be asked may seem limitless, there are limits to which aspects of topics we can study and how we can study them.

One of the most important questions in feasibility is whether or not you have access to the people you want to study. For example, let's say you wanted to better understand students who engaged in self-harm behaviors in middle school. That is a topic of social importance, to be sure. But if you were a principal in charge of a middle school, would you want the parents to hear in the news about students engaging in self-harm at your school? Building a working relationship with the principal and the school administration will be a complicated task, but it will be necessary to gain access to the population you need to study. As we discussed in **Chapter 2: Using Sociology in Everyday Life**, research must often satisfy multiple **stakeholders**—that is, individuals or groups who have an interest in the outcome of the study. Your goal of answering your research question can be realized only when you account for the goals of the other stakeholders.

School administrators also want to help their students who are struggling with self-harm, so they may support your research project. But they may also need to avoid scandal and panic, providing support to students without making the problem worse.

Assuming you can gain approval to conduct research with the population that most interests you, do you know if that population will let you in? Researchers like Barrie Thorne (1993), who study the behaviors of children, sometimes face this dilemma. In the course of her work, Thorne has studied how children teach each other gender norms. She also studied how adults "gender" children, but we'll focus on just the former aspect of her work. Thorne had to figure out how to study the interactions of elementary school children when they probably would not accept her as one of their own. They were also unlikely to be able to read and complete a written questionnaire. Since she could not join them or ask them to read and write on a written questionnaire, Thorne's solution was to watch the children. Although this seems like a reasonable solution to the problem of not being able to actually enroll in elementary school herself, there is always the possibility that Thorne's observations differed from what they might have been had she been able to actually join a class. What this means is that a researcher's identity, in this case Thorne's age, might limit (or enhance) her ability to study a topic in the way that she most wishes to study it.

In addition to personal characteristics, there are also the very practical matters of time and money that shape what you are able to study or how you are able to study it. In terms of time, your personal time frame for conducting research may be the semesters during which you are taking your methods courses or working on a thesis. Future employers will give you specific deadlines for completing research tasks. Those timelines will shape the sort of research you are able to conduct. Surveys can be completed in minutes online, but recruiting a representative sample for that survey might entail much more work. Immersive qualitative work—such as embedding yourself in an organization—can take months or years. Money, as always, is also relevant. Obtaining commercial datasets can be expensive, and finding participants willing to sit for long in-depth interviews often requires you to pay them.

Mapping the Social World: A Q&A with Jennifer A. Johnson



An expert in social network analysis (often abbreviated as SNA), Jennifer A. Johnson is a professor of sociology at Virginia Tech and the chair of its Department of Sociology. She received her PhD in sociology from the University of Virginia and later joined the faculty of Virginia Commonwealth University's Department of Sociology, which she chaired as well. Johnson also worked for the Department of Defense (DoD) as a social science analyst, receiving a Chairman of the Joint Chiefs of Staff Distinguished Civilian Service Award in 2006. Her current work analyzes the relationship between pornography consumption and sexual health and victimization among early adolescents. In another line of research, funded with a National Science Foundation ADVANCE-IT grant, Johnson and her colleagues are using SNA to explore how intersectional identities affect outcomes for women within STEM and non-STEM academic departments. In previous work, she used SNA to map how the pornography industry markets itself and connects with users. Her work has appeared in the

International Journal of Media and Cultural Politics, *Journal of Women's Health*, *Journal of Sex and Marital Therapy*, and *Archives of Sexual Behavior*.

What first drew you into sociology, and how did you decide to become a professional sociologist?

I arrived at Radford University in the 1980s as an accounting major. However, I didn't do so well in business calculus, and that let me know that maybe accounting was not for me. Then I took this class called Sex Roles. We would now call this class the Sociology of Sex and Gender. It really transformed the way I saw the world. I had feelings about myself, my parents' divorce, and my place in the world that I couldn't articulate. And that class helped me put into words all the experiences that I'd had up to that time as a young woman. It gave me the language to understand myself and the world around me in ways that deeply resonated with me.

I got my bachelor's degree, master's degree, and finally my PhD in sociology. All the while, I continued to do the same kind of core scholarship, which was to understand gender and its impact on women's lives—as a system of marginalization, oppression, and control, but also empowerment and emancipation.

How did you first come to use social network analysis in your research?

After I got my master's degree, I did a lot of community college teaching for about 10 years, but wanted to try my hand at some kind of professional practice. I landed a position with the Department of Defense at the Joint Warfare Analysis Center as a social science researcher. Social network analysis was the methodology they were using to track terrorist networks. When I went into the DoD, I had no specific knowledge of social network analysis or terrorism. What I did have was a social scientific approach to the practice of research, data collection, and analysis. I also had skill in the presentation of social science research because of my teaching background.

What are some of the ways you have used social network analysis in your professional career?

At the DoD, I used it to map, measure, and develop interdiction strategies for terrorist networks across the world. But I realized after several years that I missed teaching. I missed students. And I did not particularly like to work in an environment that was so heavy on hierarchy, very male-oriented, very masculine in its language and practices.

I obtained a position as an assistant professor at Virginia Commonwealth University, where I also served as a faculty representative for Students for Social Action. This student group invited Gail Dines to come and talk about her research on the pornography industry. At that presentation, she mentioned that we really didn't have an understanding of how mainstream businesses connected to businesses involved in the pornography industry. I realized that that would be a perfect research project for me. I could study gender systems of oppression and marginalization with my newly acquired skills.

Since then, I have used the concepts and methods of social network analysis to understand the business—the political economy—of the online pornography industry. And it's led me into many different areas. For example, I was involved in developing the digital sociology concentration in the sociology master's program at VCU, which was grounded in the appreciation my research had given me for digital economies. I'm now working to understand the impact of these economies on adolescent sexual health. I just gave a presentation to the Israeli government on the ways in which digital economies capture adolescent sexual development at the early stages and then go on to influence later stages—particularly for their primary targets, young heterosexual men. I'm also funded by a grant from Gail Dines's organization to study child-on-child sexual abuse. All the work I'm doing now stems from that connection I developed between social network analysis and digital economies.

What can social network analysis tell sociologists about the workings of social groups that other research methods cannot?

Well, when we talk about social relationships, social connections, social interactions, we are talking about what takes place between people. All of the ways in which people relate and connect to one another are about exchange: ideas, gifts, love, touch, working together. The foundation of sociological theory is the social construction of reality through interactions between and among people. Social network analysis is the visual representation and measurement of those varieties of interactions. It's based in matrix algebra and the idea that you have "nodes" and "edges." Nodes are the dots on a network diagram, and edges are the lines between the nodes.

SNA can be applied to almost any type of social unit that makes exchanges. The nodes could be computers, and the computers could operate between each other. The nodes could be businesses, and businesses could share information. In my research on the digital pornography industry, the nodes were company websites, and the edges were the links that websites shared to other websites. So, while SNA depicts social interactions, and typically the nodes are thought of as people, there are all sorts of ways to apply it.

Sociologists often consider SNA both theory and method. It's about developing a theory of how people work inside groups. How big is a group, how small is a group? How are people rewarded, how are they marginalized? What is the impact of these things on how groups function? And where are you located in the group? Are you a leader, or are you a follower? Are you at the center, or are you at the periphery? What's the impact of your memberships in different types of groups? All these questions are theoretical questions that social network analysis seeks to map and measure. It's really a descriptive methodology, although some advanced SNA is working toward making it predictive.

What are some opportunities for using social network analysis in the real world?

SNA is often used in organizations. I've used it to help nonprofits understand how they can establish better connections. Which nonprofits are more central? Which ones need better outreach? I've seen it used to make domestic violence response systems more effective. How do shelters work with emergency rooms or with police? You can look at how those systems work together and look at them against particular outcomes. SNA is also used in classrooms to understand how group work transpires. Teachers look at patterns of interactions among students to see if some patterns produce stronger learning outcomes than other patterns.

I'm currently using social network analysis to map out informal and formal interactions between faculty members in academic departments. The goal is to see if STEM departments work differently than social science or humanities departments, with a particular focus on issues relating to women. Are women in STEM departments centrally located or more peripheral? Do they do more work on committees? And I'm not referring just to numbers. Who's having lunch with whom? Who's socially marginalized, and who's not?

I've also worked with a research group to use social network analysis to tie together disparate data for the Richmond Police Department. They had two groups that were beefing, and they couldn't figure out why. They gave us a download of data, and we did a social network analysis of it that found the two groups were beefing over a girl who was a node in the network connecting these groups. Though we had no subject-matter expertise on the nature of the problem, we were able to identify this odd connection in the network.

What are some of the ways social network analysis can contribute to digital sociology?

I think that social network analysis should become a foundational methodology in digital sociology programs because so much of the digital world is about networks. Close your eyes and conceptualize what the World Wide Web looks like. It is, in fact, one large network of nodes and connections—links—between them. Standard statistical analyses often involve examination of independent variables and dependent variables. But traditional social science variables, such as the race, class, and gender of individuals, cannot easily be collected when you are looking at data produced on web networks, which is termed "native data." Researchers have to start

thinking differently about how to deal with such large amounts of native data without the presence of easily identified social categories attached to individuals.

We need theoretical and methodological development in this area. We know that students are interested in thinking critically about the impact of social media beyond how many hours you spend on social media, and whether you are depressed or not. There is a whole social world that is built through how social media platforms relate not just to their users, but also to each other. And a lot of it is hidden in the ways in which websites talk to one another, and the ways in which businesses are owned. All these things—which are the dominant forces shaping our world today—need to be studied.

What do you think are some of the challenges to the future use of social network analysis in sociology?

I think the biggest challenge is to ramp up the social network analysis skills of sociologists. I know that there's lots of interest, but there aren't many people who are well-trained in it. Some of that might be because of the need for more sophisticated computers. In network data, each data point is really three points of data—two nodes and a link between them. Social network analysis can scale up to encompass very large networks, but you need computers able to process that data.

Also, there's a need to build stronger bridges with computer scientists, mathematicians, and media specialists. Oftentimes sociologists think about their interdisciplinary colleagues as naturally being located in the humanities or other social sciences. But in digital sociology, you need an understanding of how code works. You need an understanding of how websites are built and connected to one another. You need an understanding of how to collect, or "scrape," that data. SNA datasets are different from, say, a General Social Survey dataset or others with columns of data containing variables that you can easily pull and extract. For SNA, you might have 50 columns that all have to stay aligned with one another, and that requires knowing how to manipulate data, how to manage very complicated but interrelated datasets, and how to scale up computer processes. It's the idea of thinking about how websites interact, and how websites are built with the intention of relating to particular audiences—in other words, the political economy of websites.

How do you recommend students learn more about how to do social network analysis?

There are actually lots of online training options. Search engines will find SNA analysis tools, web-based courses, and AI tools. For example, ICPSR has basic orientations for SNA.[1] NodeXL is a very easy SNA add-in for Excel. And many universities subscribe to learning services, such as Lynda or LinkedIn Learning, which have modules on social network analysis. The trick is in picking the software—it's like choosing between [statistical analysis programs like] SPSS or SAS. There are a variety of programs available, some of which are free, and you pick the one you like. In SNA, obtaining estimates is a relatively simple methodology to understand once you have the data in the correct format. The challenge is in the data collection—basically, measuring the number of people you're connected to and where you're located in the patterns of those interactions.

I am really excited by the potential for digital sociology and social network analysis, the methodology most applicable to studying the web. Young sociologists are at the cusp of a new revolution in the discipline. It's very similar to how I was in the 1980s with the sociology of sex and gender. I needed language to help me understand what I was feeling. Students need a language today to understand how they're interacting with the world. And to me, sociology is the field where they should find it. It is in the best place to marry the critical analysis and advanced methods needed to understand digital society.

1. ICPSR is the Inter-university Consortium for Political and Social Research. The organization often offers summer courses in SNA.

Writing the Sections of a Research Proposal

Once you have decided on a general methodological approach that is appropriate and feasible, you will want to start describing your planned study in a *research proposal*. Think of your research proposal as a literature review plus sections devoted to methods, hypotheses, and limitations. We have already described what goes into the literature review, but note that your review of past studies also allows you to generate hypotheses regarding what results your research is likely to uncover. The proposal's methods section, in turn, will describe in depth the procedures you will use to answer your question. A proposal often concludes with a discussion of possible implications and limitations of the proposed research. In other words, the research proposal is basically a draft version of your final empirical paper—minus the all-important results section, and including a preliminary discussion of implications that you are certain to revise in the final paper. We will cover each of these sections in your proposal.

Introduction. As we noted earlier, the introduction to your research proposal looks the same as an introduction for a stand-alone literature review. However, in both the abstract and the parts of introduction that summarize the proposed study, you will have more details to provide. We recommend the following structure for this summary:

> Past studies have failed to examine [*research problem*]. Using [*methods/data*], this proposed study examines [*research question*]. The findings of this study are expected to be [*hypotheses*], meaning that we need to rethink [*implications*].

Literature review. The literature review follows the proposal's introduction, and looks exactly as we described in **Section 5.1**.

Methods. The methods section is a plan for what exactly you will do in your study. The more specific it is, the better. Of course, at this planning stage, many of the decisions you make will be somewhat arbitrary. For instance, you may not know exactly how many people you will be able to recruit for the in-depth interviews or surveys you want to do. It may be hard even to guess how large a sample is reasonable. Nevertheless, put down some numbers, however preliminary. Those details can always change as you talk through the project with other people and figure out what is feasible.

Let's break down what goes into the methods section of your proposal. (Although some of the terms we mention might be unfamiliar to you at this stage, when you return to these guidelines after reading the following chapters, they will be much clearer!) Overall, this section of the proposal should provide a thorough discussion of the methods, data, and measures you will use to test your hypotheses, generate theory, and otherwise conduct your research. Justify your choice of methodological techniques, sampling procedures, interview and survey questions, and so on, referring to any strengths and weaknesses in the methods you plan to use, given what your research problem and research question are. There are always shortcomings or compromises involved with any proposed sampling procedure, operationalization of a theoretical concept, or other methodological choice, so you should have plenty of things to discuss in this section.

For all studies, you should discuss the proposed target population, sample size, time frame of data collection, sampling techniques, and process for recruiting any respondents. You should also review the key measures that you are focusing on, describing exactly how the underlying concepts of interest were or will be operationalized in terms of actual question wording and other specifics. Finally, discuss your analysis strategy (e.g., what statistical procedures will be used, or how qualitative data will be coded). For a proposed survey study, make sure to provide substantial detail about how the survey was or will be fielded, discussing how representative the sample is of the population of interest (including any weighting or other adjustments) and identifying the dataset and sponsoring organization (if using a secondary dataset). In a subsection called "Measures," discuss the department variables and then independent variables.

For a proposed qualitative study, include a subsection called "Case Selection" in which you elaborate on the purposive sampling strategy used in choosing sites and types of respondents during data collection. Discuss the specific inclusion and exclusion criteria for your sample, and be specific about the process for recruiting interviewees or gaining access

to sites. For studies based on in-depth interviews or surveys, you should create a full interview guide or survey questionnaire to accompany your proposal, in addition to describing your key variables in the methods section or measures subsection in as much detail as possible.

Hypotheses. For quantitative studies in particular, you will want to include hypotheses—likely answers to your research question—in your proposal, which typically follow the methods section (but can also appear in the literature review, usually at its end). For qualitative studies, sometimes the hypotheses are merely implied, and if they are mentioned, they may not be identified as "hypotheses" per se—just listed as expectations in the literature review or elsewhere.

A study's hypotheses should be derived from the scientific literature or based on logic. After you decide on the study's overall methodology and conduct a literature review, you should be able to apply what you know from the literature to the sample or context you're studying, specifying what results you expect your research to uncover.

As scientists, we test hypotheses using data. This process is called **inference**. We infer—conclude based on evidence—whether or not a hypothesis is valid. When the data don't support our hypothesis, it is rejected—or *falsified*. As we discussed in earlier chapters, your hypotheses must be **falsifiable**—that is, you must be able to confirm or disconfirm them with data obtained from your target population.

Hypotheses can be strong—in that they propose a *specific* and *causal* relationship—or they can be weak—in that they just say the two variables are (somehow) associated. **Causal inference**—testing whether changes in one variable truly cause changes in another variable—is a heavier lift than just testing whether the two variables are correlated. Consider these different hypotheses:

> *Weak*: Children's family incomes are related to their later educational attainment. (This tells us nothing about the direction of the hypothesis—whether the relationship is positive or negative—or its causality—whether greater family income actually causes higher educational attainment.)
>
> *Stronger*: Children's family incomes are *positively* related to their later educational attainment. (This hypothesis indicates the directionality but not the causality of the relationship.)
>
> *Strongest*: Children's family incomes have positive effects on their later educational attainment. (This hypothesis specifies both the directionality and the causality—that greater family income causes higher educational attainment.)

Scientists typically want to test strong hypotheses because they want to know the directionality and causality of a particular relationship. That said, sometimes the data available won't allow you to test those aspects of a relationship; as we'll discuss in **Chapter 12: Experiments**, certain research designs are better at evaluating whether a relationship is truly causal. Regardless of whether our hypotheses are weak or strong, however, they should clearly specify independent and dependent variables. In the hypothesis, "children's family incomes have positive effects on their later educational attainment," it is clear that income is the independent variable (the "cause") and educational attainment is the dependent variable (the "effect"). It is also clear that this hypothesis can be evaluated as either true (if higher family income raises later educational attainment) or false (if higher family income has no effect on, or lowers, later educational attainment).

Note that hypotheses can be simply described within the literature review or provided in a separate section. Sometimes, they are listed formally as hypothesis 1, hypothesis 2, and so on. You should follow the lead of similar studies found in your literature review, given that different fields and subfields have different norms for presenting hypotheses.

Conclusion and limitations. There isn't much to say in the conclusion to your research proposal, since you haven't yet conducted your study. Thus, this section will necessarily be short. That said, you should be able to broadly discuss the possible implications of your study for research, theory, practice, or policy, giving particular attention to how your study connects to the existing literature. You should circle back to the research problem that you argued for at the end of your literature review or in a separate "Statement of the Problem" section, making the case for how your study will resolve

that problem. If the limitations of your study were not fully discussed in the methods section, you should point them out, given the relevant methodological and theoretical approach.

Key Takeaways

1. The research design you choose should follow from the research question you ask.
2. Feasibility is always a factor when deciding what, where, when, and how to conduct research. Aspects of your own identity may play a role in determining what you can and cannot investigate, as will the availability of resources, such as time and money.

5.7. References

Booth, Wayne C., Gregory G. Colomb, Joseph M. Williams, Joseph Bizup, and William T. FitzGerald. 2016. *The Craft of Research.* 4th ed. Chicago: University of Chicago Press.

Laubepin, Frederique. 2013. *How to Read (and Understand) a Social Science Journal Article*. Ann Arbor, MI: Inter--university Consortium for Political and Social Research (ICPSR).

Lurie, Shira. 2014. "How to Read a Book in Two Hours or Less." *Inside Higher Ed*, January 14.

Munn, Zachary, Micah D. J. Peters, Cindy Stern, Catalin Tufanaru, Alexa McArthur, and Edoardo Aromataris. 2018. "Systematic Review or Scoping Review? Guidance for Authors When Choosing between a Systematic or Scoping Review Approach." *BMC Medical Research Methodology* 18(1):143. doi:10.1186/s12874-018-0611-x.

Thorne, Barrie. 1993. *Gender Play: Girls and Boys in School.* New Brunswick, NJ: Rutgers University Press.

White, Lynn. 2005. "Writes of Passage: Writing an Empirical Journal Article." *Journal of Marriage and Family* 67(4):791–98. doi: 10.1111/j.1741-3737.2005.00175.x.

6. SAMPLING



How can a survey of a small group of people tell us what the entire population of the United States thinks about a particular topic? To accurately measure public opinion, researchers do not necessarily have to talk to everyone in the population of interest. Rather, they can talk to a **sample**, a group of people who *represent* the entire population. This approach is similar to when a doctor examines a few drops of blood (i.e., a blood *sample*) and draws conclusions about the constitution of a person's blood throughout their whole body.

Sociologists are usually interested in making claims about a larger group of people—what we have called a **population**. A population could be all adults in a country, or all workers at a company, or all students at a university, but it could also be a particular group—say, college athletes, international students, or older adults living with diabetes. It could also be something other than a group of *people*, such as social media posts published on one platform in the past year, horror movies released in the 1990s, protest events during the civil rights era, and so on. Even if the population that sociologists want to learn about is relatively small, it's usually impossible for them to study all of the members or parts of that group. So, they typically choose to study a sample—that is, a segment of the group that contains a manageable number of **cases** (units from the population). The cases in the sample are supposed to represent the larger population. What a sociologist learns about that particular sample will hopefully be true for the entire population.



When we conduct research, we rarely have a chance to interview or observe every member of the population we're interested in studying. Instead, we look at a subset, or sample, of that population, which hopefully serves as a good stand-in for the whole. *Gerd Altmann, via Pixabay*

Sampling has to do with this process of selecting a subset of the population of interest and using it to draw reasonable conclusions about the characteristics of that population. Proper sampling is critical when conducting many kinds of empirical research. For example, biologists collect their data by examining and measuring a sample of the organic matter they're studying, such as plants or soil. In a similar way, sociologists obtain samples of *social* information, often by interviewing or observing groups of people.

In this chapter, we will learn about various approaches to sampling and the reasons that sociologists might use certain techniques while avoiding others. We will talk about how we can judge the quality of particular samples and sampling strategies, and what problems can occur when we obtain skewed samples that don't accurately represent what is going on in the broader social world. In the process, we will identify basic questions that should be asked whenever a study uses a sample to make claims about a population.

How researchers sample, and *whom* or *what* they sample, fundamentally shapes the conclusions that they are able to draw from their data. That said, we want to stress that there are many appropriate ways to conduct sampling for a given research project. Like other aspects of sociological research, sampling is an art in addition to a science, and some of the best work arises when sociologists think creatively about how and whom to sample.

Opening chapter image credit: *Karolina Kaboompics*, *via Pexels*. *Adapted by Bizhan Khodabandeh.*

6.1. Units of Analysis



Before developing a sampling strategy, researchers must first define who or what is the target (the unit of analysis) of their scientific study. *Rory McKeever, via Unsplash*

Learning Objectives

1. Describe units of analysis.
2. Discuss how we can study the same topic using different units of analysis.

Before you can decide on a sampling strategy, you must define the unit of analysis of your scientific study. The **unit of analysis** refers to the person, collective, or object that you are focusing on and want to learn about through your research. As depicted in **Figure 6.1**, your unit of analysis would be the *type* of entity (say, an individual) you're interested in. Your sample would be a *group* of such entities (say, a group of individuals you survey)—which, collectively, stand in for the *population* you wish to study.

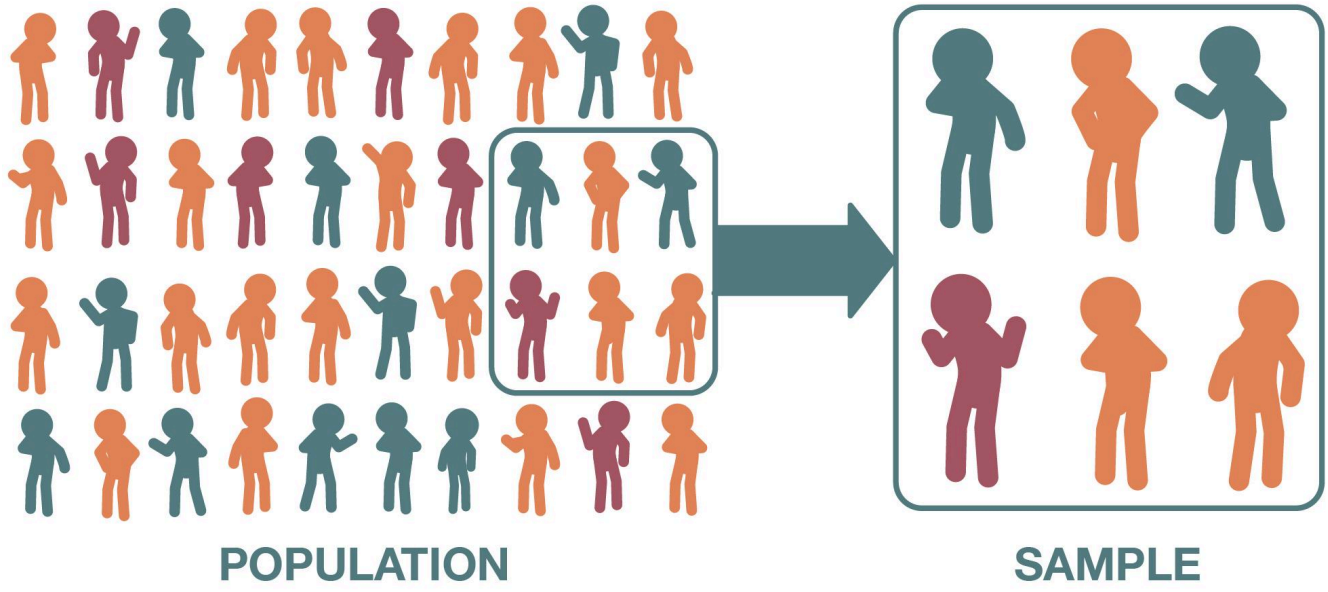


Figure 6.1. Units of Analysis, Populations and Samples. When conducting research, we focus on studying a particular *unit of analysis*—a person, collective, or object that we want to learn more about. That unit of analysis encompasses a larger *population* we are interested in making claims about. It is usually not possible to include every unit of the population in a research study, so we must select a small subset of the population (a *sample*) from which we actually gather data.

Typical units of analysis include individuals, groups, organizations, and countries. For instance, if we are interested in studying people's shopping behavior, their learning outcomes, or their attitudes toward new technologies, then the unit of analysis is likely to be the *individual*. If we want to study characteristics of street gangs or teamwork in organizations, then the unit of analysis is probably the *group*. If our research is directed at understanding differences in national cultures, then our unit of analysis could be the *country*. In the latter two examples, even though specific individuals—the group or country's leaders—may have a greater say over what these groups or countries do, for the sake of analysis, researchers typically think of those decisions as reflecting a *collective* decision rather than any one individual's decision.

Even inanimate objects can serve as units of analysis. For instance, if we wish to study how two or more individuals engage with each other during social interactions, the unit of analysis might be each *conversation*, and not the individual speakers. If we wanted to track how depictions of people of color have changed in popular culture over time, we could focus on a *film* or *television show* as a unit of analysis.

Our choice of a particular unit of analysis will depend on our research question. For instance, if we wish to study why certain neighborhoods have high crime rates, then our unit of analysis becomes the *neighborhood*—not crimes or criminals committing such crimes—because the object of our inquiry is the *neighborhood* and not the people living in it. If, however, we wish to compare the prevalence of different types of crimes—homicide versus robbery versus assault,

for example—across neighborhoods, our unit of analysis could very well be the *crime*. If we wish to study why criminals engage in illegal activities, then the unit of analysis becomes the *individual* (i.e., the criminal).

Now let's consider a completely different kind of sociological study. If we want to examine why some business innovations are more successful than others, then our unit of analysis is an *innovation*—such as the invention of a new method for charging phones. If, however, we wish to study how some tech companies develop innovative products more consistently than others, then the unit of analysis is the *organization*. As you can see, two related research questions within the same study may have entirely different units of analysis.

Determining the appropriate unit of analysis is important because it influences what type of data you should collect for your study and whom you collect it from. If your unit of analysis is the organization, then you usually will want to collect *organizational-level* data—that is, data that has to do with the organization, such as its size, personnel structure, or revenues. Data may come from a variety of sources, such as financial records or surveys of directors or executives, who are presumed to be representing their organization when they answer your survey questions. Meanwhile, if your unit of analysis is a website, you will want to collect data about different sites, such as how one kind of site compares to others in terms of traffic. We could use the term "site-level" data—just like we'd use the term "individual-level" data when individuals are the unit of analysis. We could also talk about "lower" and "higher" levels of analysis—with individual-level data existing on a lower level than group-level data, which may, in turn, be on a lower level than national data (see the discussion of *micro*, *meso*, and *macro* levels of analysis in **Chapter 3: The Role of Theory in Research**). It is important to note that "higher" does not imply "better" in this case. We're just talking about whether we're looking at smaller or larger groupings of data.

Frequently, the unit of analysis is what we observe in our research—the source of our data—but that is not always the case. In fact, sometimes we want to make a distinction between units of analysis and **units of observation**. The unit of analysis is what we really want to study, but sometimes we have to get at it indirectly, by observing something else. For example, surveys often ask questions about families to understand their family structure, income, and various aspects of their well-being, but they need to get information about the family through individuals—specifically, the respondent who is answering survey questions on behalf of the family. In this case, the unit of analysis for the survey's family-related questions would be the family, but the unit of observation would be the individual. Likewise, in our earlier examples, we talked about studying organizations and websites as our units of analysis, but doing so might involve talking to individuals—the directors of those organizations, or the users of those websites, respectively.

Analyzing multiple types of units of observation can give us a fuller picture of our unit of analysis. For example, if you are conducting research about what makes particular social media apps more addictive than others, then examining differences between the apps in terms of their functionality (*app* as the unit of observation) would tell you one thing, but surveying individuals about their usage of apps (*user* as the unit of observation) would clarify other aspects of that question. Furthermore, it is often a good idea to collect data from a lower level of analysis and sum up, or *aggregate*, that data, converting it into higher-level data. This can give you a bigger-picture perspective on your unit of analysis. For instance, to study teamwork in organizations, you can survey individuals in different teams and measure how much conflict or cohesion they perceive on their teams. You can then average their individual scores to create a "team-level" score on those particular ratings. Note, however, that issues can arise when we move in the opposite direction—from a higher to a lower level of analysis (see the sidebar **Deeper Dive: Ecological Fallacies**).

Ultimately, the unit of analysis will help you determine both the population you are interested in and the sample that you will study to arrive at any conclusions about that population. So you need to choose it wisely. For example, let's say you're interested in the average pay of chief executive officers (CEOs) at companies across the nation. The unit of analysis would be the CEO, and the population would be all individuals in the country who work as company CEOs. But the unit of analysis would be different for a very similar research question: the average amount that U.S. companies pay their CEOs. In this case, the unit of analysis is actually the *company* because you are interested in how much companies pay their CEOs—not how much individuals are paid as CEOs. The difference is subtle, but the main point is that your

unit of analysis is linked to whatever population you actually want to say something about—in this example, either individual CEOs, or companies that have CEOs.

Deeper Dive: Ecological Fallacies



When researchers confuse their units of analysis and observation, they may commit an *ecological fallacy*—that is, making possibly inaccurate claims about individuals based on aggregated data collected at a higher level. *lil artsy, via Pexels*

A mismatch between the unit of analysis and the unit of observation can create issues for researchers. Let's say you want to compare the residents of different states (your unit of analysis is the *individual*), but you only have access to state-level data (your unit of observation is the *state*). This is a problem because you generally do not want to be making claims about a lower level of analysis based only on aggregated data at a higher level—in this example, drawing conclusions about individuals based on the states where they reside. For instance, the fact that the population of a state is, on average, wealthier than the rest of the country does not mean that residents of that state are more likely to be rich than the average American. It may be that a small contingent of superrich people have pulled up the average wealth of the state, but its many other residents actually tend to be poorer than the average American. (As you might know from your statistics classes, in this situation, *mean* wealth—the group's average—differs dramatically from *median* wealth—how much money the person smack in the middle of the income distribution has.) This logical error—making claims about the nature of individuals based on data from the groups they belong to—is called an **ecological fallacy**.

Émile Durkheim's classic study of suicide is often mentioned as an example of an ecological fallacy. One of the pioneers of the field of sociology, Durkheim argued in his 1897 book *Suicide* that societies in which individuals struggled to feel they belonged—that is, populations with low levels of social integration—would experience more suicide. Ideally, the unit of analysis for such a study would be the individual. Specifically, we would want to study individuals and the factors that contributed to their deaths by suicide. But Durkheim did not have individual-level data. Instead, he had higher-level data about the number of suicides in each country. To test his theory that social integration safeguarded individuals against suicide, Durkheim compared countries that were mostly Protestant to those that were mostly Catholic. The idea was that Protestantism was a more individualistic and unstructured faith than Catholicism, and so the two varieties of religious belief could stand in for less and more social integration, respectively.

Durkheim's analysis concluded that suicide was indeed higher in Protestant-majority countries. The problem

was that his data only allowed him to say that Protestant *countries* were more likely to have higher suicide rates—not that Protestant *individuals* were more likely to commit suicide. To conclude the latter would have been an ecological fallacy, and yet that was the question that Durkheim truly wanted to answer. To his credit, Durkheim also tested his theory by studying suicide rates across localities *within* countries—another level of analysis (Selvin 1958). (Replicating your analysis across different types of data is a good way to check the robustness of your findings, as we will discuss in later chapters.) Durkheim found the same pattern of higher levels of Protestant belief correlating with higher suicide rates within counties, giving further credence to his theory. Although flawed, Durkheim's analysis made creative use of the data that was available to him at the time, and his work continues to inspire researchers, including those studying the growing rates of suicide among less educated Americans since 2000 (Case and Deaton 2020).

Key Takeaways

1. A unit of analysis is a member of the larger group you wish to be able to say something about at the end of your study. A unit of observation is a member of the population that you actually observe.
2. When researchers confuse their units of analysis and observation, they may commit an ecological fallacy—that is, when we make possibly inaccurate claims about the nature of individuals based on data from the groups they belong to.

6.2. Populations versus Samples

Learning Objective

Understand the difference between populations and samples.

Deciding on your unit of analysis will help you specify what your **population of interest**, or **target population**, is—the larger group that your research question applies to. Populations in research may be rather large and encompassing, such as "people who live in the United States" or even "all human beings" (indeed, for many medical or psychological studies, the latter is the implied target population). The populations we study as sociologists are typically less vague, however. For example, even a large study that purports to study the views of "all Americans" will in practice specify which "Americans"—adults, citizens, legal residents, voters, or some combination thereof—they can or wish to study.

Now that you've identified your unit of analysis and your population of interest, you can start interviewing and observing folks, right? Not so fast. You want to think long and hard about your strategy for obtaining your sample, that subset of your population of interest from which you'll actually gather data. This systematic and intentional approach to sampling is part of what distinguishes *scientific* research from other kinds of inquiry. Much of the rest of this chapter will be focused on best practices for obtaining scientific samples.

You might be protesting at this point—why do we need to bother with a sample? Why can't we just talk to everyone in our population of interest? It is actually rare for researchers to gather data from an entire population. This might sound surprising or disappointing until you think about the kinds of research questions that sociologists typically ask. For example, let's say we wish to answer the following research question: "How do the middle school experiences of students of various racial and ethnic backgrounds differ, and how are they similar?" Would you be able to collect data from all students across all middle schools in the country? In truth, you will rarely have the time or resources needed to gather data from every single person your study might be interested in. This means you have to make hard choices about sampling and then be honest with yourself (and with the readers of the paper you plan to write) about the limitations of your sample.



If you were trying to understand the opinions of a large group—say, the preferences of a stadium crowd about a team's new mascot—you would be foolish to try to survey every person in that target population. Instead, you would want to take a subset of that group—a *sample*—and ask those people your questions. If your sample is properly chosen (as we will describe, *randomly selected*), then even a sample of dozens of people could give you fairly accurate estimates of what the entire group thinks. *Anna Sullivan, via Unsplash*

Both qualitative and quantitative researchers use sampling techniques to help them identify what or whom they will actually study. Because the goals of qualitative and quantitative research differ, however, so, too, do their sampling procedures. The two broad categories of sampling strategies are **probability sampling** (also called **random sampling**) and **nonprobability sampling**. Probability sampling is widely used in quantitative studies, and nonprobability sampling is more common for qualitative work.

Probability sampling helps us make precise claims about populations much larger than the actual samples that we gather. This approach is ideal if we want a high degree of confidence that the results of our study are **generalizable**—that they apply to the broader population from which our sample was drawn. Researchers (especially quantitative researchers) typically want their findings to be generalizable, but nonprobability sampling can also be justified under certain circumstances. The point of this nonprobability approach is often to sample in a way that allows researchers to closely examine the workings of a particular phenomenon and flesh out their theoretical understanding of it.

Quantitative and qualitative researchers also have different criteria for judging whether a sample is "sufficiently large"—in academic lingo, whether it has a high enough n, or number of cases.[1] We'll go into those criteria next, but it's worth recognizing that the most significant factor influencing the sample sizes of both qualitative and quantitative

1. Typically, a lowercase n refers to the number of units in a sample, and an uppercase N refers to the number of units in the population; however, some authorities, such as the American Medical Association, use other conventions, such as treating N as the size of the sample and n as the size of subgroups within that sample.

research is often the study's budget. Larger samples provide us with greater confidence in our results, but they take more time, effort, and money to obtain. Conducting in-depth interviews or doing ethnographic observations is particularly labor-intensive, which is one of the reasons that qualitative studies usually have many fewer cases than quantitative studies.

Now let's break down the features of a sample. Each case in our sample—say, each individual we interview—is called a **sampling unit**. When we analyze a sample quantitatively, we generate **statistics**—numbers like means, medians, and standard deviations that describe the sample (or some group within our sample). For example, if we collected a sample of the U.S. adult population, we might want to calculate statistics like the mean level of education of all the respondents within our sample, or the standard deviation of the annual incomes of any respondents who work for pay. Statistics are our estimates of what the actual numbers are for the population. Those population **parameters**—the means, medians, standard deviations, and so on for the whole population—are the numbers we *really* care about but that we usually cannot observe (that is the whole point of sampling). For instance, whenever we generate a statistic about support for a particular presidential candidate based on a political poll's sample of surveyed voters, we are hoping that statistic tells us not only the percentage of the candidate's supporters *in that sample* but also the percentage in the *overall* population of voters. With proper sampling, our statistic should be a good estimate of its corresponding population parameter. When our sample's statistics generally match up well with the population parameters, we say our sample does a good job "representing" the larger population of interest. It is a **representative sample**, and any conclusions we draw from that sample should be generalizable to the population.

Throughout this chapter, we will be talking about populations and samples of *people*. Remember, however, that your object of interest as a researcher—your unit of analysis—does not necessarily have to be people. Some sociologists study songs, news stories, textbooks, government policies and laws, movies, and other symbolic or physical materials. (We'll talk about specific strategies for studying some of these units of analysis in **Chapter 15: Materials-Based Methods**.) For example, in a study of media representations of immigration in Williamsburg, Virginia, sociologists Deenesh Sohoni and Jennifer Bickham Mendez (2014) used texts that referenced immigration as their sample. Specifically, they collected and analyzed news stories, letters to the editors, op-ed pieces, columns, and public commentary published in a local newspaper, the *Virginia Gazette*, in 2006 and 2007. These published articles made up their sample of cases to analyze. Note that their choice of a local newspaper to study is itself a sample—a sample (albeit of one case) meant to stand in for the media more broadly. (We'll say more about sampling decisions at different levels later in this chapter when we discuss purposive sampling, which often relies on such an approach.)

This brings us to another bit of general advice about sampling: don't be afraid to be unconventional and creative in choosing units of analysis and target populations to study or in identifying particular samples to gather. You don't have to stick with the obvious approach of interviewing or surveying a group of people who have opinions on your topic. Many of the sampling methods described in this chapter can be used to select cases for human-based and materials-based research alike.

Representativeness: Matching Populations and Samples



When choosing our sample, we must ensure that it matches the population we are interested in studying. In other words, our sample must be *representative* of the target population. Representativeness means that your sample contains the same sorts of variation that are present in your larger population of interest. If, for example, you are interested in learning about differences between men and women, you should ensure that your sample doesn't contain only men. Similarly, if your target population includes individuals with diverse demographic characteristics (e.g., a wide range of ages, races, and socioeconomic statuses), you want to ensure that many of those groups are represented in your sample. *Austin Distel, via Unsplash*

As we've discussed, social scientists draw samples from the populations they're interested in studying. Sometimes, however, you may find that the population you are drawing your sample from isn't the population you're *really* interested

in. For instance, many psychological studies are truly interested in how human beings think and act—not just how Americans think and act, or how college students think and act. Yet if you've taken any psychology classes, you've probably noticed that the lab experiments that are the foundation of empirical work in this field largely recruit college students as their subjects—hungry undergrads who sign up to take lab tests in exchange for, say, $10 or a free donut. One analysis found that more than two-thirds of participants in papers published by top academic journals in psychology were located in the United States, and two-thirds of studies using U.S. samples that appeared in one well-respected journal, the *Journal of Personality and Social Psychology*, obtained their data from undergraduates taking psychology courses (Arnett 2008).

What are we actually learning from such studies? Joseph Henrich and his colleagues (2010) point out that behavioral scientists often make sweeping claims about human nature based on samples drawn only from **WEIRD societies** (Western, educated, industrialized, rich, and democratic). In fact, many widely accepted findings about human behavior—in regards to fairness, cooperation, visual perception, and trust, among other things—are based on studies that excluded participants from outside the United States and sometimes excluded anyone outside the college classroom (Begley 2010). This raises questions about what we really know about *human* behavior, as opposed to U.S. resident behavior or U.S. undergraduate behavior.

Even if a lab experiment was conducted on U.S. psychology undergrads, isn't it still likely to tell us something meaningful about how human beings as a whole think and act? Yes, behaviors relating to cooperation and trust might vary greatly across cultural contexts, but college students from rich democracies aren't all that different physiologically and cognitively from the rest of humanity, are they? The problem is that we will not know for sure until we do more research with non-WEIRD samples. This doesn't mean we should dismiss those earlier studies, but we do need to be aware of their limitations. (This issue is similar to the one we previously discussed concerning units of observation and analysis, when the population we observe is not the same as the population we truly care about.) Although the theoretical conclusions of these studies can be generalized to the population that their samples were drawn from (say, U.S. psychology undergrads), it is not so clear that they also apply to the real population of interest (say, human beings).

Another issue that will come up frequently in this chapter is that of *representativeness*. As we discussed earlier, a representative sample captures important characteristics of the population from which it was drawn, in ways that are critical for the accuracy of the research being conducted. For example, if you wish to report about differences between men and women at the end of your study, your sample should obviously have sufficient numbers of both of these genders. In other words, your sample should contain the key sorts of variation that are present in your larger target population.

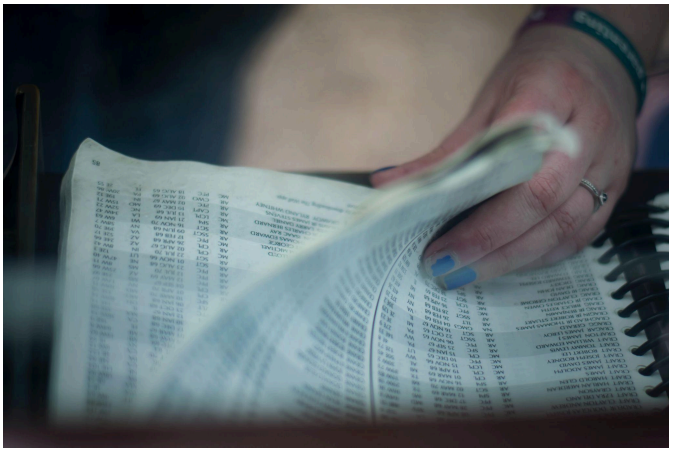
These sorts of concerns about a sample's representativeness plague all social scientific research methods, not just lab experiments. Your choice of a particular sample will always depend on how feasible it is for you to collect it. The decision may also rest on how strong a justification you can make for your sample's generalizability. For example, suppose you want to survey people about their preferences for cilantro—that notoriously divisive herb. Although some international surveys include respondents from many nations, most social scientists have the resources to study, at best, only the people in their country. And even if you could do a national study of cilantro, do you really have to? Do we expect the cilantro preferences of Virginians to be different from those of Arizonans? Probably not. If you expect no differences, then you might be justified in gathering a sample from a smaller, state-level population, with the assumption that the preferences of Virginians are similar to those of Arizonans. This assumption would not be reasonable, however, if you were studying attitudes about the federal government, because the political views of Virginians differ from those of Arizonans in notable ways. In that case, you would probably want to sample the entire U.S. population, or sample certain groups within that population that you think would be theoretically relevant.

As we described in **Chapter 3: The Role of Theory in Research**, every theory we generate has a set of **scope conditions** (explicit or implicit) that tell us where it applies, and where it does not necessarily apply. The extent to which our

population of observation matches our population of interest provides us with a clear set of boundaries that constrain the generalizability of our findings. These limitations should be expected for any study.

Indeed, we can and should ask tough questions about the WEIRD-ness of the samples often used in the social sciences, but we must stress that all research is hard to do, and all studies have weaknesses. Any decisions we make, including decisions about sampling, inevitably have trade-offs. After all, sociologists who use other methods look with envy at lab researchers who have easy access to a whole bunch of (presumably) willing and able human guinea pigs to recruit for their samples. That access has allowed fields like psychology (including the sociological subfield of social psychology) to flourish, even if it has arguably come at the cost of sample representativeness. The more practical lesson to be drawn is that we should pay close attention to how well the population we are observing fits the population we are really interested in. We should be upfront about any mismatch and talk about our results with these sorts of limitations in mind. And we should ask hard questions about other studies that generalize their findings beyond the population they were actually able to sample.

Sampling Frames and Sampling Bias



A *sampling frame* is a list of people (or elements) from which you will draw your sample. Putting together a sampling frame is typically the first step in conducting human subjects research. Researchers must think about locations or groups in which their target population gathers or interacts. For example, let's say that qualitative researchers interested in understanding the quality of the care provided in nursing homes have chosen to study a local nursing home. They could obtain a sampling frame by asking the home's staff to provide a list of all their patients. They could then select participants for the study from that list. If they eventually decided to include more nursing homes in their study, then the sampling frame could expand to cover all patients at all the selected nursing homes. *Brittany Colette, via Unsplash*

Another task that researchers need to complete before sampling can begin is finding—or creating—a sampling frame. A **sampling frame** is the list of every single member, or **element**, of your population of interest. Sometimes, the sampling frame is more or less obvious. If your population of interest is all the students at a school, then presumably the school's administration has that list somewhere in their records. But what if the administration is not willing to share that list with you (a common occurrence)? Well, then you might decide to use your school's public directory as your sampling frame. As we will describe later, this sort of approach has flaws, too, but at least we have some way of generating relevant names.

What about sampling a huge and sprawling population—like all U.S. residents? A publicly accessible list of all Americans (residents or citizens) is not available to draw upon. Researchers, however, can get lists of all U.S. addresses. This is the sampling frame that many government surveys and other large-scale studies use to obtain nationally representative samples. Mailings can be sent to a sample of those addresses, with surveyors making follow-up visits (like those of census enumerators) as needed.

Regardless of the approach, sampling frames rarely match the population of interest perfectly, as depicted in **Figure 6.2**. For instance, if you used your school's public directory as a sampling frame to identify its student population, clearly you would be missing certain people who chose not to be included in the directory or were left out for some other reason. Even if you had the resources to do mailings to a sample of all U.S. addresses, this approach would still exclude many Americans—most obviously, the homeless. In these two

cases, we are dealing with **sampling bias** (sometimes called *selection bias* or *sample selection bias*).[2] **Bias** is a systematic error that may make our research findings inaccurate in some way (we'll talk more about *systematic* versus *random* error in the next chapter). When this bias arises as a result of the sample selection approach we are using, we call it sampling bias.

Bias in sampling occurs when the elements selected for inclusion in a study do not do a good job of representing the larger population from which they were drawn. In the examples we mentioned, certain important segments of the population are "missing" in our sample—that is, they are underrepresented compared with their proportion in the actual population. This means that our sample is not representative, and the statistics we generate will be off the mark ("biased" or "skewed") in comparison to the population parameters.

Later, we'll say more about the different ways that bias can creep into samples, but it's worth reiterating that researchers are not always so lucky to have access to the complete lists that organizations, corporations, or government agencies maintain of their employees, members, and citizens. And sometimes no list exists, period. Say that you wish to conduct a study of hairstyle preferences across the United States. Just imagine trying to create a list of every single person with hair in the country. When faced with scenarios like this one, social scientists might construct their own sampling frames. In the world of political polling, for instance, random-digit dialing has been a popular—if problematic—way of addressing the fact that no obvious list exists of American voters to contact. The assumption is that every modern American adult has a phone, so by calling numbers randomly, you should be able to reach every American adult. As assumptions go, this is not too much of a stretch, and political polls have operated reasonably well in spite of the obvious sampling bias that exists in this scenario—among other things, the fact that Americans who do not own phones are excluded from the sample.

Casting such a wide net with your sampling frame means you have to be especially diligent in weeding out all the cases in your generated list that do not fit your population of interest. For instance, if you are trying to survey American voters, then you don't want to talk to anyone who picks up the phone who isn't an American citizen and thus cannot vote. These cases would fall into the "Ineligible" category shown in **Figure 6.2**. Including these respondents in your survey results would create sampling bias, because your sample wouldn't faithfully reflect the underlying population.

2. Although selection bias and sampling bias are often used interchangeably, some scientists distinguish the two terms, with sampling bias being a subset of selection bias.



Figure 6.2. Target Populations and Sampling Frames. Any sociological research project includes a *target population* (of people, collectives, or objects) that you are interested in learning about. Because we normally cannot study every member of that population, we need to gather a sample, and to obtain that sample, we need to find or generate a *sampling frame*—a list of all the elements in the target population. Note that the sampled population does not overlap with the actual population—obviously, because it is a subset of that population, but also because of sampling bias. The sampling frame may also capture cases that do not fit within the target population, and that therefore must be excluded from selection during the sampling process.

Take another look at **Figure 6.2**. Notably, the boxes for both the sampled population and the sampling frame overlap but do not match the box for the target population (note that the diagram is not drawn with any set proportions in mind). Obviously, our sample is usually much smaller than the population of interest, but it may be biased as well—that is, it may not reflect the characteristics of the true population of interest because of some systematic problem. The sampling frame we use may create some of that bias—for example, if it underrepresents certain subgroups of the population, or if it includes ineligible cases that we failed to weed out.

You may have noticed that in **Figure 6.2** we labeled the middle box "Sampled Population" rather than "Sample." We want to point out a key distinction, which is linked to another key type of sampling bias. For a survey, the **sampled population** is all the people we sought to recruit from our population of interest. But it is almost never the case that everyone selected for the sample actually responds to the survey. Some people may have died or moved away, and others may decline to participate because they are too busy, are not interested in the survey topic, or do not participate in surveys on principle.

Take a look at **Figure 6.3**. Our sampled population is now divided into two groups. Those who actually take our survey are called **respondents**. These are the people whose responses are used to calculate our statistics, and this is our actual *sample*, whose size we report as n in our papers. Those who did not take the survey for whatever reason are called **nonrespondents**. Their numbers are frequently not reported in papers but can be inferred from our **response rate**—the percentage of all the people we asked who did take our survey (which is n divided by the number of people in the sampled population).

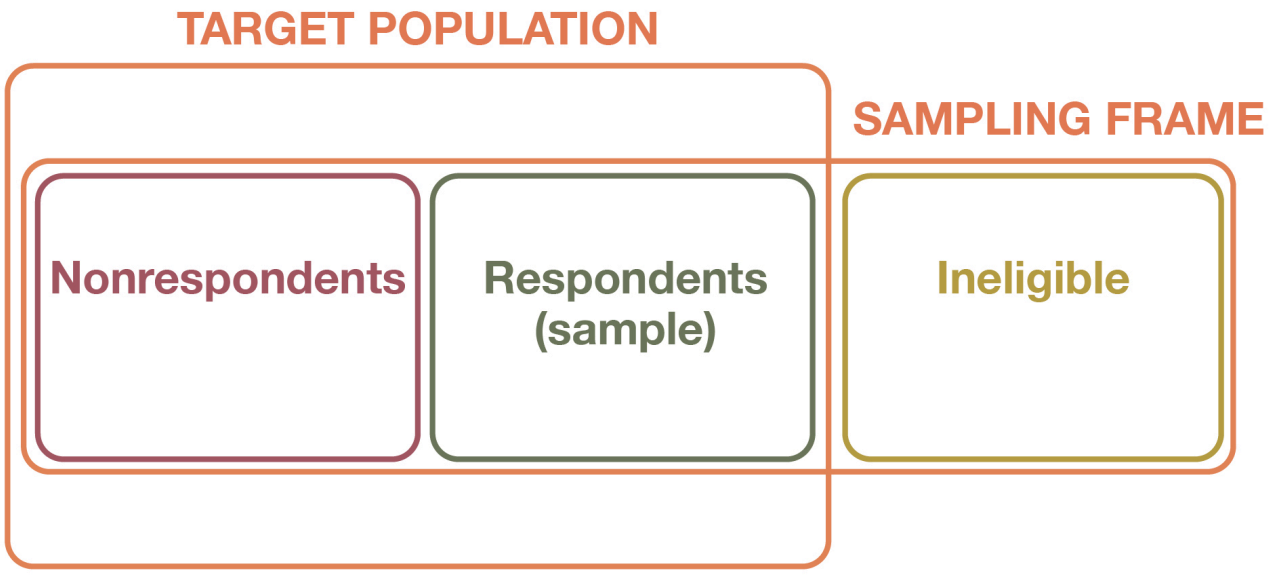


Figure 6.3. Respondents versus Nonrespondents. If we are conducting a survey, we will usually fail to recruit some of the potential respondents in our sampling frame: we won't be able to reach them, or they will refuse to participate. These individuals are referred to as *nonrespondents*. If they look different from our respondents in ways that are important to our study, then our sample will suffer from *nonresponse bias*.

A low response rate is not a problem in itself, but it is generally a worrisome signal. As we will describe in **Chapter 13: Surveys**, modern-day political polls are able to provide competent estimates of voter opinion even with astoundingly (and increasingly) low response rates. The more important consideration is whether survey nonrespondents differ from survey responders in systematic ways. If there is a difference, then we may be dealing with **nonresponse bias**. For example, if Republicans are less likely to answer questions for a political poll (a real concern today in U.S. polling), then our sample will be skewed toward Democrats, and our sample will not be representative of actual U.S. voters. Response rates do figure into this equation because the lower the response rate, the more reason we have to believe that respondents and nonrespondents look different on key characteristics.

We'll say more later in the chapter about the methodological issues that nonresponse bias can cause. For now, just be aware of the many ways that our samples can fall short of capturing the real population of interest.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://viva.pressbooks.pub/sociology-research-methods/?p=197#oembed-1

Video 6.1. How a Flawed Sample Sank a Major Presidential Poll. Biased sampling led to a significant error in the *Literary Digest*'s predictions of the 1936 U.S. presidential election. Although the sample size was huge, the sampling frame did not reflect the population. This led to fundamentally flawed predictions and an embarrassing situation for *Literary Digest*. For that same election, the polling company Gallup was able to

correctly predict the winner of the election using a substantially smaller sample. This case shows how correct sampling procedures can be more important than sample size.

Deeper Dive: Sampling Gone Wrong

When fielding a typical survey, it is extremely important to choose a sample that truly captures the population of interest so that any conclusions drawn from the sample can be generalized to the population of interest. Improper and biased sampling is a key reason that political polls sometimes get the outcomes of elections very wrong.

Video 6.1 illustrates how a biased sampling strategy gave way to a massive prediction error in the 1936 presidential election between incumbent president Franklin D. Roosevelt and Kansas governor Alf Landon. *Literary Digest*, one of the country's most popular magazines at the time, had correctly predicted the presidential elections of 1920, 1924, 1928, and 1932 by sending postcards to a sample of Americans and asking whom they were planning to vote for. In 1936, they were confident the same strategy would work. Not only had *Literary Digest* been right for four straight presidential elections, but no one could beat their sample sizes. In 1936, they sent out 10 *million* postcards, getting back more than 2 million responses. (In political polls today, pollsters feel good about getting a *thousand* respondents.) Bigger is always better, right?

As it turned out, no. Even though the sample size was huge, the *Literary Digest* poll utilized a flawed sampling frame, which drew from lists that the magazine was able to obtain of people who owned automobiles and telephones. In 1936—the depths of the Great Recession—automobiles or telephones were amenities that only more affluent households could afford. As a result, the sample left out many poorer Americans—who that year broke in large numbers for the Democratic candidate, Roosevelt (Brockell 2020). Another major problem—arguably the more decisive factor (Lusinchi 2012; Squire 1988)—was nonresponse bias. The people who were likely to bother responding to a postcard survey and mailing it back were voters with strong opinions. At the time, Roosevelt was a polarizing figure who drew fierce criticism for his New Deal policies, which conservatives likened to socialism. It was no surprise, then, that the respondents who sent in their postcards disproportionately favored Roosevelt's opponent.

Because the poll's respondents did not adequately reflect the country's electorate, *Literary Digest* blew its presidential prediction that year. The magazine declared that Landon would win 31 states. He ended up winning only two. Roosevelt snatched up an unprecedented 62 percent of the national vote, even though *Literary Digest* concluded he would win only 43 percent, a stunning prediction error of 19 percentage points. Meanwhile, an upstart polling outfit founded just a year earlier by journalist George Gallup conducted its own survey of voters, gathering a sample of 50,000 respondents—still an unnecessarily large number by modern standards. Gallup used a new technique called *quota sampling* (described later in the chapter) to generate a sample that more closely mirrored the actual U.S. electorate. Although Gallup's sample was a tiny fraction of *Literary Digest*'s, he correctly predicted that Roosevelt would win. Adding to the humiliation, Gallup took *Literary Digest*'s sampling frame and drew his own sample of just 3,000 respondents, which, in a test of his new statistical methods, he used to precisely predict the magazine's polling results before they were released (Brockell 2020).

Literary Digest folded just two years after its election debacle. Its massively incorrect prediction based on an astounding sample size of two million respondents has taught many generations of social scientists since then

an important lesson: the size of a sample is not as important as its representativeness. You are better off with a small but well-chosen sample than a large but biased one.

Key Takeaways

1. A population is the group that is the main focus of a researcher's interest; a sample is the group from which the researcher actually collects data.
2. Populations and samples might be one and the same, but more often they are not because we generally do not have the time or resources to survey or examine every member of a population of interest.
3. Sampling involves selecting the observations that you will analyze, and it always opens up the possibility that your sample may not be representative of the target population because of various kinds of sampling bias or just bad luck.

Exercises

1. Read the methods sections of a couple of scholarly articles describing empirical research. How do the authors talk about their populations and samples, if at all? According to their methods sections, which specific target populations are the researchers drawing conclusions about?
2. Think of a research project you have envisioned conducting as you read this text. Would your population and sample be one and the same, or would they differ somehow? Explain.

6.3. Probability Sampling



Probability sampling refers to sampling techniques that share one key condition: we know the likelihood that a person (or unit of analysis) will be selected for inclusion in our sample. Knowing this likelihood helps us identify a representative sample from which we can collect our data, ensuring that whatever results we find are generalizable to the population we are actually interested in. *Patrick Fore, via Unsplash*

Learning Objectives

1. Describe how probability sampling achieves generalizability.
2. Identify the various types of probability samples, and understand the logic behind each approach.

All probability sampling techniques have two attributes in common: (1) every single unit in the population has a chance (a nonzero probability) of being selected for the sample, and (2) the sampling procedure involves a random process of selection at some point. We'll break down what this means in a bit, but the first thing to know is that the goal of this sampling approach is to achieve generalizability. This is the key distinction between probability samples and nonprobability samples. As we alluded to earlier, **generalizability** refers to the idea that the results obtained from a given sample apply to the population from which it was drawn. If this is the case, then we say our study's results are "generalizable" to the population of interest.

This brings us to the first of our two conditions: when we use probability sampling, every person (or other unit of analysis) in the population should in theory be able to be selected for the sample, and we should know the likelihood of that selection. Why do we care about this likelihood that a particular case will be selected? Knowing this information allows us to make sure our sample is representative—that it does not look substantially different from the population as a whole. Having a representative sample means, in turn, that we can generalize any statistics we derive from that sample back to the population. If our sample is truly representative, then we can be confident that any means, standard deviations, and other figures that we calculate from it also apply to the larger population.

Note that a representative sample resembles the population from which it was drawn in *all the ways that are important* for the research being conducted. If, for example, your study intends to say something about differences between men and women, you better make sure that your sample doesn't contain only women. That's a bit of an oversimplification, but the point with representativeness is that if your population varies in some way that is critical to your study, your sample should contain the same sorts of variation.

We mentioned earlier that researchers should know the likelihood that each case in their population of interest will be included in their sample. What does this mean? For the more advanced sampling techniques we'll describe later, the probability of being selected into a sample can actually vary, but for the moment, we can think of this likelihood of selection in simpler terms: all possible cases in the researcher's target population have an equal chance of being selected for inclusion in the study. To ensure this outcome, we rely on **random selection.** In practice, this means picking cases from the population at random—like flipping a coin, or drawing numbers from a hat—so that every individual has an equal chance to participate in our survey, every news article has an equal chance to be included in our content analysis, and so on. If the chance of being selected for a sample is equal across all the cases in a given population, then according to statistical theory, the sample should be representative of that population.

Some of the probability sampling techniques that sociologists frequently use are *simple random* sampling, *systematic* sampling, *stratified* sampling, and *cluster* sampling. We will go over each strategy, but first we want to say something about two critical topics in probability sampling: sampling error and sample sizes.

Using random selection does not mean that your sample will be perfect. No sample is perfect. The only way to obtain a perfect result—that is, a precise estimate of the actual values in the population of interest—would be to include everyone in the population in your sample, which defeats the whole point of sampling. Generalizing from a sample to a population always contains some degree of **error**—that is, some amount of inaccuracy in your measurement of the actual values in the population. When researchers use random selection techniques to draw a sample, however, they are actually able to estimate how closely the sample represents the larger population from which it was drawn. Specifically, they are able to calculate the **sampling error**—that is, the difference expected between the results obtained from a sample and the actual parameters of a population. That is an incredibly powerful feature of probability sampling, which gives studies using it a high degree of precision and credibility—assuming, of course, that sampling bias has not skewed the sample.

In quantitative research, the size of the sample is critical because it affects the sampling error. Larger samples are subject to less sampling error—the larger the sample, the closer any statistic will be to the actual value in the population, again assuming your sample is truly random. In fact, quantitative researchers can use statistical theory to calculate the size of the sample they need to estimate a value in the population of interest and can be fairly confident that their estimate closely matches the actual value. Technically speaking, quantitative researchers do not need to obtain a single "right" number of cases for their samples, given that the ideal number depends on this specific level of confidence and the precision they desire for their results.

That said, in sociology, you typically see an n of at least 100 for a statistical analysis to be published. For reputable national surveys, you usually see sample sizes of at least 1,000, although political polls—which will often be fielded numerous times in the run-up to an election—may just have a few hundred respondents. Higher-profile surveys in need of greater precision and coverage go further: for example, the General Social Survey recruits 2,000 or more respondents

when it fields its questionnaires every other year, and the Current Population Survey, a monthly survey conducted by the U.S. Census Bureau that is used to calculate the federal government's official employment statistics, has a sample size of 60,000 households.

Simple Random Sampling



Doing simple random sampling is like rolling dice. For every six-sided die we roll, the six outcomes—one, two, three, four, five, or six—are equally likely to occur. In a similar fashion, every member of the target population ideally has the same likelihood to be selected for our sample when we are using simple random sampling. *Guillermo Velarde, via Unsplash*

Simple random sampling is an intuitive way of generating a representative sample, given how random selection will ensure that the characteristics of that sample reflect those of the larger population. In simple random sampling, all members of the population (more accurately, of the sampling frame) are given an equal probability of being selected. In other words, each case is identified strictly based on chance.

To draw a simple random sample, you must start with a sampling frame—again, a list of every single element, or member, of your population of interest. Number each element in that list sequentially and then randomly select elements. Then collect data from those elements. For example, if you were conducting a survey of students at your university, you might obtain a list of students from the university's administration, give each student a unique number, and then use a random number generator to pick students by their numbers. (Websites like Stat Trek and Randomizer.org provide random

number generators that you can use for samples of whatever size.) You would contact those students and continue to recruit respondents until you had enough people in your sample.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://viva.pressbooks.pub/sociology-research-methods/?p=205#oembed-1

Video 6.2. Random Sampling. This video from the Pew Research Center provides a brief explanation of random sampling and why it's important.

Systematic Sampling

Systematic sampling techniques (illustrated in **Figure 6.4**) are somewhat less tedious but offer the benefits of a random sample. As with simple random samples, you must be able to produce a sampling frame, a list of every person in your population. Rather than having to use a random number generator, however, in systematic sampling you simply select elements from your list in specified intervals—say, every *k*th element on your list. The number you use in place of *k* is your **selection interval**, or the distance between the elements you select for inclusion in your study. To begin the selection process, you first need to figure out how many elements you wish to include in your sample. Let's say you want to interview 25 fraternity members on your campus, and 100 people on campus are members of fraternities. To define your selection interval, simply divide the total number of population elements by your desired sample size. In this case, you would divide 100 by 25, so *k* would be four, and you would want to select every fourth element in your list. You do not, however, just start from the first element in the list. To allow for an equal chance that every element could be selected, it is important that the starting point be randomly chosen from within the first *k* elements on the list. For instance, in our fraternity study, we would randomly choose from among the first, second, third, or fourth name on our list as our first potential interviewee to contact. Systematic sampling would then involve the selection of every fourth name from that point in the list onward.

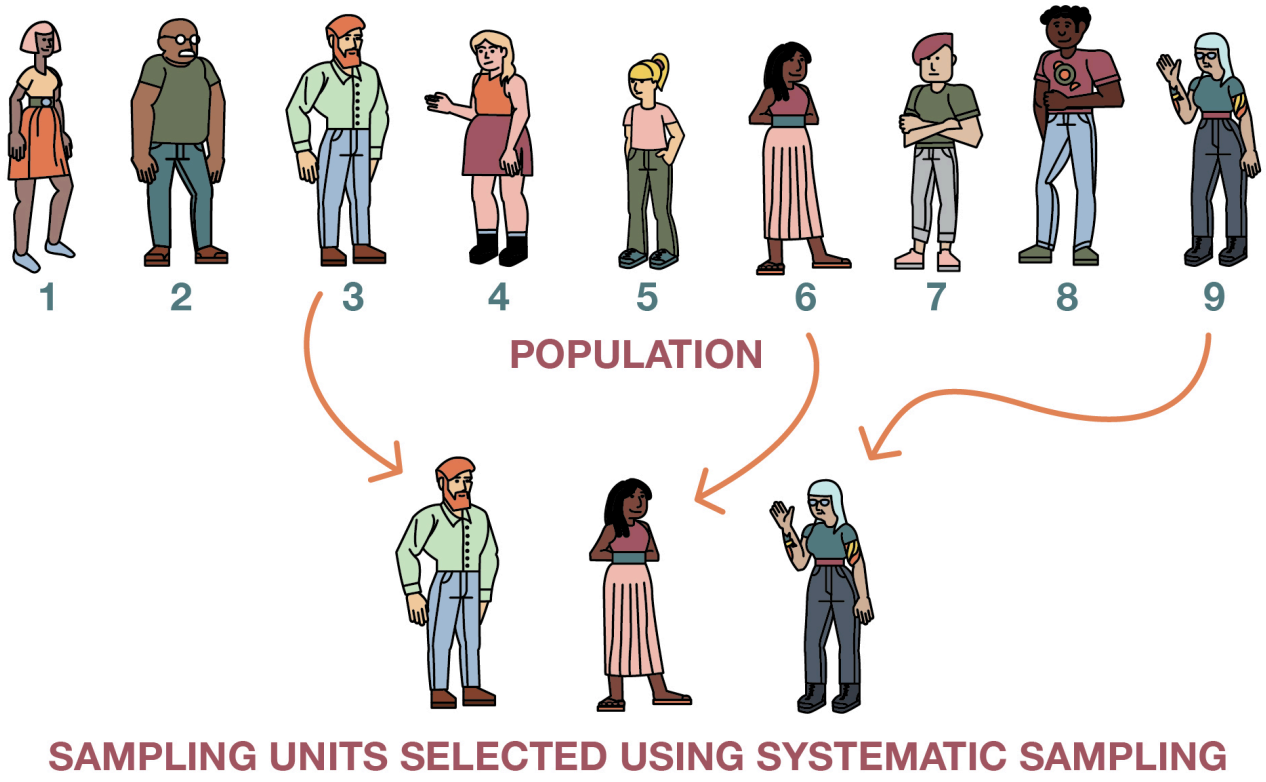


Figure 6.4. Systematic Sampling. Systematic sampling involves systematically choosing sampling units from a list by defining an interval (*k*) and then selecting every *k*th element in our list. In this example, the interval is three, so we selected every third individual from our population. Note, though, that to generate a truly random sample—in which every element in your sampling frame has a chance of being selected—you should randomly choose your first case before you start selecting every *k*th case. In this example, we chose the third person and began counting off from there, but our sampling process should have allowed for us to start with the first or second person, too.

There is one scenario where systematic sampling should clearly not be employed. If your sampling frame has any pattern to it, you could inadvertently introduce bias into your sample by using systematic sampling. This is sometimes referred to as the problem of **periodicity**—the tendency for a pattern to occur at regular intervals. For instance, let's say we wanted to randomly select four days in a month during which we plan to conduct observations at the campus library. We decide to use systematic sampling to identify our observation days, following the procedure we've already outlined:

We divide our population size (30 days) by our desired sample size (four days), arriving at a selection interval of seven.

1. Out of the first seven cases, we randomly select the second one as our starting point.
2. Given our selection interval of seven, we select every seventh day after our starting point.
3. You can see how this procedure works out (or, rather, doesn't work out) in **Figure 6.5**:



Figure 6.5. An Example of Periodicity When Using Systematic Sampling. Systematic sampling is not the best strategy to use when our sampling frame presents a pattern at regular intervals. In this example of a calendar, weekends occur at regular intervals, so a systematic sample of days might pose problems if the selection interval was specifically seven (though not at other selection intervals).

Do you notice any problems with our selection of observation days in Figure 6.5? Apparently, we'll only be observing on Tuesdays. As you have probably figured out, that is not a good plan if we really wish to understand how public spaces on campus are used. Weekend use probably differs from weekday use, and that use may even vary during the week, just as class schedules do.

Stratified Sampling

In cases such as the one presented in Figure 6.5, in which the sampling frame is cyclical, it would be better to use stratified sampling. In **stratified sampling**, a researcher will divide the study population into two or more mutually exclusive subgroups (also known as **strata**) and then draw a sample from each subgroup. In this example, we might wish to first divide our sampling frame into two lists: weekend days and weekdays. Once we have our two lists, we can then apply either simple random or systematic sampling techniques to each subgroup.

Stratified sampling is a good technique to use when a subgroup of interest makes up a relatively small proportion of the overall sample. For example, in our observation of the campus library, we want to be sure to include both weekdays and weekends in our sample. Because weekends make up less than one-third of an entire week, it is possible that a simple random or systematic strategy would not yield sufficient weekend observation days. To ensure that we have sufficient coverage, stratified sampling is the way to go.

In fact, stratified sampling becomes essential when we want to make sure that a tiny subgroup is adequately represented in our sample. For example, suppose we are conducting a survey of people's perceptions about the police in Richmond, Virginia. Clearly, it would be important to include the opinions of members of many different racial and ethnic groups in such a survey, given how critical these factors are thought to be in shaping how the police interact with civilians. But some racial and ethnic groups make up a very small percentage of the population in Richmond. According to the U.S. Census Bureau's American Community Survey, the city's population in 2018 was 46.2 percent African American, 41.5 percent white, 7.1 percent Hispanic, 2.4 percent Asian, and 2.8 percent other. As a result, simple random or systematic sampling might not yield many—or even *any*—respondents from those groups, depending on how large our overall sample is.

By using stratified sampling, however, we can ensure that our sample contains sufficient numbers of such individuals. In this case, our *strata* (the segments into which our sample is divided) are racial and ethnic groups. We would carve up our sampling frame into these strata, and then choose a number of cases from each stratum—each racial and ethnic group—in line with its numbers in the larger population. For instance, if we were stratifying by race in our Richmond sample, 46.2 percent of our respondents should be African American, 41.5 percent should be white, and so on. (Obviously, we can round the exact proportions to match what is possible in our smaller sample.)

The process we have just described is **proportionate stratified sampling** (depicted on the left side of Figure 6.6):

the sample sizes for our subgroups match their size in the population. Sometimes, though, we want to sample even more people from particular subgroups—a greater number than their proportion within the population. For example, in our survey of Richmonders' perceptions of the police, a sample of 100 people would give us just two or three Asian respondents under proportionate stratified sampling, given that Asians account for 2.4 percent of the city's population. Even a sample of 1,000 Richmonders would give us just 24 Asian respondents, which is lower than what we need to obtain reasonably precise statistics for a group. (A rough rule of thumb is that you need at least 30 cases to calculate meaningful statistics for a particular sample or subsample.)

In this case, we might decide to pursue **disproportionate stratified sampling** (depicted on the right side of **Figure 6.6**), which would mean that the proportions of each subgroup within the sample would not reflect those in the sampling frame (or the population of interest). Specifically, one or more smaller subgroups would be **oversampled**—they would end up representing a greater share of the sample than they do in the larger population. As we saw with our example of Asian respondents in Richmond, disproportionate stratified sampling can be indispensable when we are interested in learning about small subgroups because it allows us to obtain a sufficient sample size for those groups. The downside of this approach is that it messes up our statistics for the overall sample. Now we have "too many" cases from the subgroups we oversampled, and the means, standard deviations, and other numbers from our sample are no longer generalizable to the population of interest. In making sure our sample better represented particular subgroups, we inadvertently made it less than representative of the entire population.

Fortunately, we can correct for this issue by **weighting** our data. Weighting adjusts how much a particular case contributes to the statistics for a sample. The technical procedure for weighting is beyond the scope of this textbook, but you can think about it, in metaphorical terms, as making some cases heavier—making them matter more in any calculations—and making other cases lighter—making them matter less. Suppose we oversampled Asian respondents so that our sample included *twice* as many Asians than their proportion in the overall Richmond population would call for (meaning that we wound up with, say, 48 Asian respondents when a proportionate sample would be 24). When we calculate statistics for the entire sample, we would want our Asian respondents to contribute *half* as much to those statistics—to matter half as much in our calculations—than racial and ethnic groups that are proportionately represented in our sample. Statistical programs will handle these calculations for us, and if we weight in this way, then our statistics will remain generalizable to the population of interest. (As we will describe later, weighing is not just used in disproportionate stratified sampling as a way to compensate for oversampling; it also is widely applied to improve sampling more generally, to adjust for systematic biases in the data collection, and to make the sample more representative.)

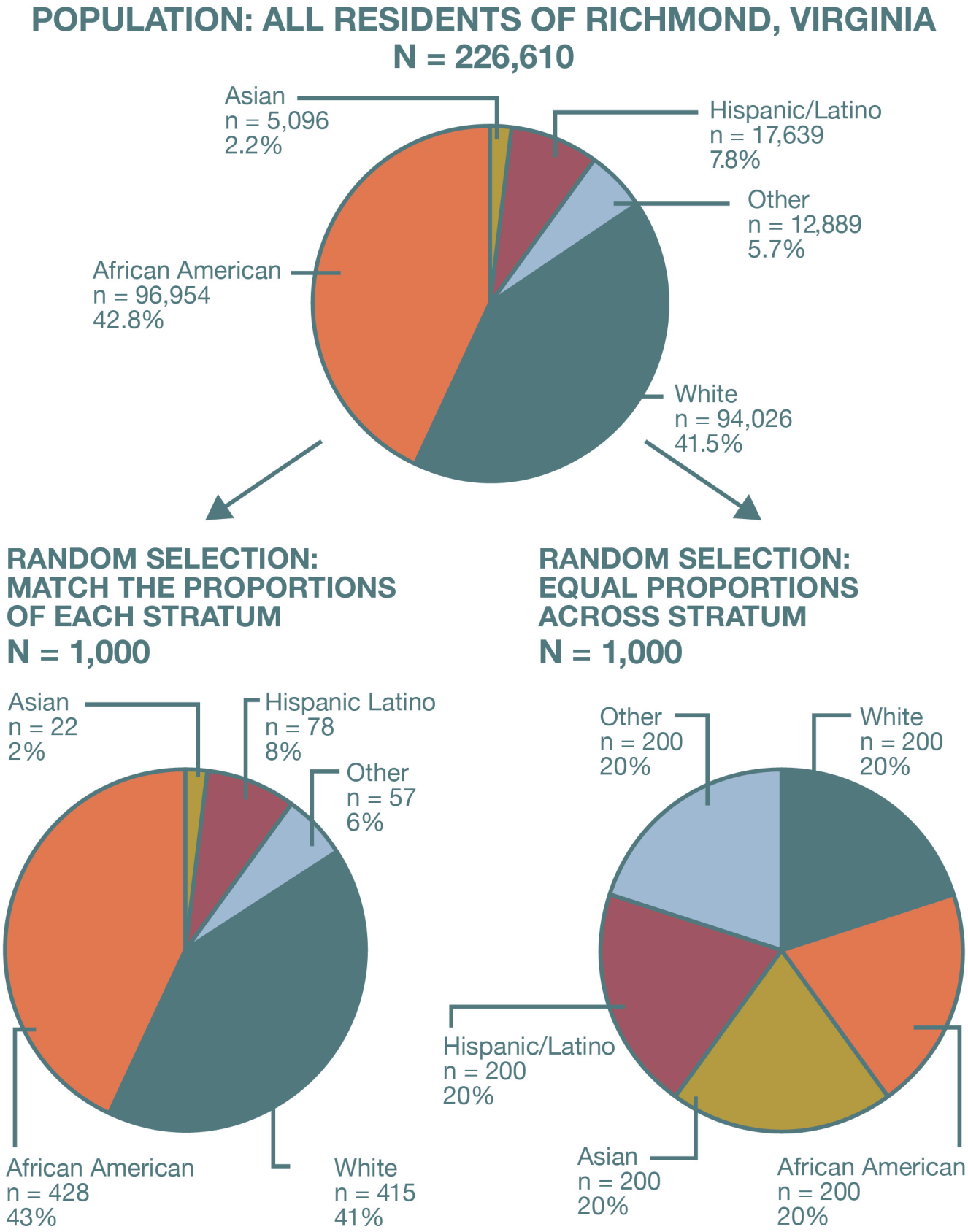

POPULATION: ALL RESIDENTS OF RICHMOND, VIRGINIA
N = 226,610
Asian
n = 5,096
2.2%
Hispanic/Latino
n = 17,639
7.8%
Other
n = 12,889
5.7%
African American
n = 96,954
42.8%
White
n = 94,026
41.5%
RANDOM SELECTION:
MATCH THE PROPORTIONS
OF EACH STRATUM
N = 1,000
Asian
n = 22
2%
Hispanic Latino
n = 78
8%
Other
n = 57
6%
African American
n = 428
43%
White
n = 415
41%
RANDOM SELECTION:
EQUAL PROPORTIONS
ACROSS STRATUM
N = 1,000
Other
n = 200
20%
White
n = 200
20%
Hispanic/Latino
n = 200
20%
Asian
n = 200
20%
African American
n = 200
20%

Figure 6.6. Examples of Proportionate and Disproportionate Stratified Sampling: Racial and Ethnic Groups in Richmond, Virginia. When we use *proportionate stratified sampling* to ensure adequate representation for particular groups in our sample, we match the proportions in the target population—finding percentages for each group in line with those in that population of interest. If some of those groups would be too small for statistical analysis, however, we might want to use *disproportionate stratified sampling*—obtaining disproportionately large numbers of cases (an *oversample*) from certain groups to allow for that analysis, while *weighting* those groups in an analysis of the overall sample so that their disproportionately large numbers don't skew our results for the total sample.

Cluster Sampling

Thus far, in our discussion of probability samples, we have assumed that researchers will be able to access a list of population elements and then select sampling units from that list. As we pointed out earlier in the chapter, however, sometimes the sampling frame does not exist or is not readily available. For example, let's say you are interested in the workplace experiences of public librarians in Virginia. Chances are that obtaining a list of all librarians who work for Virginia's public libraries would be difficult. As we noted earlier, you could generate a sampling frame using techniques like random-digit dialing—say, calling all numbers with Virginia area codes, and then talking only to those respondents who identify themselves as public librarians. Yet this would clearly not be the most practical use of your time or resources, given how many phone numbers you would need to call to gather a sufficient number of public librarians for your sample.[1]

In situations like these, researchers can turn to **cluster sampling**. A researcher using this strategy first samples groups (or **clusters**) of population elements and then selects elements from within those groups. We already noted how hard it would be to get a list of all public librarians in Virginia. But what about coming up with a list of all public libraries in the state? That should not be too hard. And once you do so, you can apply cluster sampling in two stages. First, you would draw a random sample of libraries (your cluster). Then, you would draw another random sample of elements from each cluster—librarians who work at the libraries you initially selected. To make the generation of our lists even easier, we could do our cluster sampling in the three stages shown in **Figure 6.7**—randomly selecting (1) counties in Virginia, (2) specific libraries in each county, and (3) individual librarians.

1. Of course, using Virginia area codes would also fail to reach many Virginia residents whose cell phones do not have Virginia area codes.



Figure 6.7. An Example of Cluster Sampling: Surveying Public Librarians in Virginia. If you have a population dispersed over a wide geographic region, it may not be feasible to conduct simple random sampling of the entire population. In such a case, it may be reasonable to divide the population into *clusters* (usually along geographic boundaries), randomly sample a few clusters, and measure all units within that cluster. For instance, if you wish to sample public librarians in Virginia, rather than travel across the state to interview key librarians (as you may have to do with a simple random sample), you can cluster libraries based on their counties, randomly select a set of counties, randomly select libraries within those counties, and then interview librarians who were randomly selected from those libraries. *Bizhan Khodabandeh, with a vector map via FreeVectorMaps.com*

Cluster sampling is a highly efficient method for creating sampling frames, although this approach does have some downsides. First, sampling in multiple stages introduces the possibility of greater error, as each stage is subject to its own sampling error. Furthermore, cluster sampling skews the sample in particular ways that you need to address if you want to generalize your results to the population of interest. Remember that with random sampling, the goal is to ensure that each element has the same chance of being selected. Yet using the two-stage cluster sampling approach we described in the previous example would mean that librarians at smaller libraries would be *more* likely to be selected. This is because our random sampling of the clusters—the libraries—would be just as likely to choose a small library as a large one, and yet more people tend to work at larger libraries—which would give the unfortunate librarians working there a lower chance of being selected for your study. Likewise, for the three-stage sampling approach, librarians who work in smaller counties would be more likely to be selected, given that smaller counties have just as much a chance to be selected as large counties.

How do sociologists deal with this problem? Consider a study by Jessica Holt and Wayne Gillespie (2008) that surveyed college students to understand what personal characteristics were associated with perpetrating violence in their intimate relationships. The researchers randomly selected 14 classes on their campus and then drew random samples of students from those classes. Clearly, however, not all college classes are the same size. If Holt and Gillespie had selected the same number of students from each class to complete their survey, then students in the smaller classes would have had a greater chance of being selected than students in the larger classes.

As Holt and Gillespie (2008:256) point out in the methods section of their paper, they used a "weighted random sample of classes" to correct for this problem. What does that mean? As we discussed earlier, statistical weighting generally involves adjusting samples to make them more representative of the population of interest. In cluster sampling, researchers adjust the sample selection process itself through a procedure called **probability proportionate to size** (PPS). Remember that cluster sampling with clusters of different sizes means that elements in the smaller clusters—in

this case, students in the smaller classes—have a higher chance of being selected. To ensure that each element across the clusters has an equal chance of being selected, the PPS method gives the *clusters themselves* different chances of being selected for the sample. In their study, Holt and Gillespie used this statistical procedure to calculate different chances for 14 classes to be selected for their sample. Larger classes had a higher chance of being selected, so that ultimately each student across those 14 classes had an equal chance of being part of the study.

Choosing among Probability Sampling Strategies

In **Table 6.1**, we summarize the advice we've provided about when to use each probability sampling technique and what problems might arise. As we have stressed, using a probability sample allows researchers to be on firmer ground whenever they make conclusions about larger groups, because the results from their sample can be justifiably generalized to the population of interest. To use probability sampling, researchers need to be methodical in their approach. First, they need to find or create a sampling frame. Next, they need to select elements from that sampling frame at random—for instance, drawing individuals to survey randomly from a list. This use of random selection reduces error and bias in terms of how well the sample represents the population from which it was drawn.

Table 6.1. When to Use Probability Sampling Strategies

Sampling Strategy	**Suggested Use**	**Key Problems**
Simple random sampling	When you want the most straightforward approach to random sampling	Requires a way to generate random numbers Requires a sampling frame
Systematic sampling	When you want an easy way to select cases for your random sample	May create problems if your data has a pattern (periodicity) Requires a sampling frame
Stratified sampling	When you want to make sure you have sufficient numbers of cases for small subgroups in your populations	Requires you to have detailed information about the target population Requires a sampling frame
Cluster sampling	When you don't have a sampling frame but want to randomly sample a population	Its multistage procedure creates additional sampling error

Random selection is a powerful tool for social scientists. Random selection ensures the representativeness of a sample and the generalizability of any results. It is a key feature in all probability sampling approaches, which are summarized in **Video 6.3**.



One or more interactive elements has been excluded from this version of the text. You can view them online here: https://viva.pressbooks.pub/sociology-research-methods/?p=205#oembed-2

Video 6.3. Probability Sampling Strategies. This video describes five common methods of sampling in data collection. Each has a helpful diagrammatic representation. (Note that the video also covers *convenience sampling*, which is a nonprobability sampling strategy discussed in the next section.)

We have discussed why you would use each of the probability sampling approaches above, but which one is best for your own study? In determining which strategy (or mix of strategies) makes the most sense, it helps to know more about your population. Simple random samples and systematic samples both require a list of all elements in your sampling frame. Systematic sampling is slightly easier in that you do not have to use a random number generator; instead, you use a sampling interval that is easy to calculate by hand.

Although both of these approaches are relatively straightforward, they may fail to capture sufficient cases with important characteristics in your population. Stratified samples can better account for periodicity by creating strata that reduce or eliminate its effects. Stratified samples also ensure that smaller subgroups are included in your sample, thus making your sample more representative of the overall population. These benefits are important, but creating strata also requires knowing information about your population before beginning the sampling process. In our example of conducting a survey about perceptions of the police in Richmond, we did know how many people of each racial and ethnic group live in the city, which enabled us to make sure our sample contained the same proportions. If the true population parameters are unknown, however, stratified sampling becomes significantly more challenging.

Stratified sampling is about ensuring sample representativeness *across* groups, and cluster sampling focuses on sampling *within* groups. Unlike the other three sampling methods, cluster sampling does not require you to have a list of all elements in your sampling frame. It is particularly useful for making claims about a large population, given that sampling frames are usually hard to come by for such groups. Because sampling occurs at multiple stages when you use a cluster sampling approach, your sampling error increases, and you also have to be careful to weight your sample selection so that the principle of random selection—an equal chance for every case to be selected—still holds.

Key Takeaways

1. In probability sampling, the aim is to identify a sample that resembles the population from which it was drawn.
2. The types of probability samples include simple random samples, systematic samples, stratified samples, and cluster samples.
3. Probability sampling usually requires a real list of elements in your sampling frame, whereas cluster sampling can be conducted without one.

Exercises

1. Of the four sample types described, which seems strongest to you? Which seems weakest? Explain.
2. Imagine that you are about to conduct a study of people's use of public parks. Explain how you could employ each of the probability sampling techniques described earlier to recruit a sample for your study.
3. Now let's say you wanted to do a survey of U.S. college students to understand the effect of drug use on academic performance. How would you obtain a random sample? How and why might this approach be different from the one you adopted for **Exercise 2**?

6.4. Nonprobability Sampling

Learning Objectives

1. Define nonprobability sampling, and describe instances in which a researcher might choose a nonprobability sampling technique.
2. Describe how probability sampling differs from nonprobability sampling.
3. Describe the different types of nonprobability samples.

When we engage in *nonprobability sampling*, we do not know the likelihood that a person (or other unit of analysis) will be selected for membership in our sample. Because this likelihood is unknown, we don't know whether or not our sample accurately represents the larger population of interest. That is okay, because precisely representing the population is not really the goal when we use nonprobability sampling techniques. However, we do want to have a specific purpose in mind that justifies our use of these approaches and guides the selection of our cases.



When we use nonprobability sampling, we don't know the likelihood that a person (or other unit of analysis) will be selected for membership in our sample. That means our sample may or may not represent the larger population of interest. It does not mean, however, that our sample should be drawn arbitrarily or without any specific purpose in mind—ideally, we should pursue a *purposive sampling* strategy, as described next. *Heather Ford, via Unsplash*

Under what conditions might we decide to use a nonprobability sampling strategy? One occasion would be when we are initially designing a research project. If we are conducting a survey, for instance, we may want to administer the questionnaire beforehand to a few people who seem to resemble the folks we are interested in studying to help work out kinks in the survey. Indeed, nonprobability samples are quite useful during the early stages of a research project, when we may want to conduct exploratory research to determine the scope of an issue, learn about how stakeholders define or understand a problem, and so on. Nonprobability sampling allows us to quickly gather preliminary data and get some idea of the contours of the phenomenon we are studying before embarking on a more extensive study. (By the way, we use the term **pilot testing** to describe any preliminary vetting of a survey questionnaire, interview guide, or other research instrument; **pilot study** is a broader term used for research that explores initial ideas or tests procedures as part of a more comprehensive project.)

From these examples, we can see that nonprobability sampling can be helpful for setting up and framing later research that would use more representative samples. But some sociologists and other quantitative researchers regularly turn to nonprobability sampling in conducting their research. For instance, as shown in the following video by the Pew Research Center, given how difficult it has become to conduct a representative survey of large populations like the U.S. electorate, pollsters are increasingly using a nonprobability sampling to give them a diverse sample that they can then adjust statistically to match the intended population.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://viva.pressbooks.pub/sociology-research-methods/?p=214#oembed-1

Video 6.4. Opt-In Surveys. This video from the Pew Research Center provides a brief explanation of why *opt-in surveys* (a type of survey using nonprobability sampling) are generally easier and cheaper for polling organizations to conduct. They also tackle the question of whether these kinds of surveys are accurate.

As we have previously discussed, quantitative researchers quantify social phenomena, reducing social life to numbers that allow for easy comparison and the processing of huge numbers of cases. In contrast, qualitative researchers seek to gather all the relevant details that pertain to a phenomenon—including characteristics that are hard to label or measure—with the goal of capturing its complexity and social context. They tend to work with many fewer cases given that their in-depth exploration requires greater time and effort for each one. For all these reasons, probability sampling approaches do not make as much sense for qualitative studies as they do for quantitative ones.

This view is not always appreciated. As Mario Small (2009) points out, qualitative sociologists sometimes convince themselves that to be "scientific" they somehow need to generate random samples with hundreds of cases. Yet any qualitative approach is bound to run into problems if it tries to mimic probability sampling, Small argues. For one thing, because qualitative sociologists rely on time-intensive methods of data collection, such as in-depth interviews, potential research subjects are even less willing to participate in such studies than they are in surveys. This creates low response rates and heightens the danger of nonresponse bias. Instead of pursuing this fruitless quest to achieve sample representativeness, Small suggests that qualitative sociologists should play to the strengths of their method—among other things, their ability to develop rapport with their participants and probe for more information, which allows them to gather information that is more trustworthy, detailed, and potentially surprising than what a survey could obtain. Rather than trying to obtain generalizable results about the relationships between key variables, they should focus their energies on fleshing out the specific causal mechanisms that link those variables—the specific processes that lead from one thing to another—and identifying unexpected factors and connections that survey researchers might not think to measure.

Generally speaking, qualitative sociologists rely on nonprobability sampling approaches to help them obtain an in-depth understanding of an issue rather than generalizable results. They use the data gathered from nonprobability samples to develop richer and fuller theories of social life—sometimes with the expectation that these theories will be tested later with probability samples that better represent the population of interest.

These aims of in-depth understanding and theoretical insight also guide the decisions that qualitative researchers make regarding whether or not they have gathered an adequate number of cases for their samples. Rather than hitting a set threshold of cases, qualitative researchers often seek to attain saturation—reaching a point at which collecting more data does not generate further insights into the phenomenon of interest. For an interview-based study, for example, a researcher might decide they have reached saturation and stop doing more interviews when any new interviews become less informative and do not yield any surprises. With this criterion in mind, qualitative researchers who use qualitative methods can publish a respectable study based on a small number of interviews or observations. Depending on the context, depth can be more useful than breadth. That said, most ethnographies require months of observation, and most interview-based studies include dozens of interviewees—with large-scale qualitative studies sometimes drawing from hundreds of interviews.

The various types of nonprobability samples that researchers use include the following: *convenience* samples, *snowball* samples, *quota* samples, and *purposive* samples. As we noted, quantitative researchers (especially political polling organizations) routinely use quota samples. The others are more commonly used by qualitative researchers. The last technique—**purposive sampling** (also known as *theoretical sampling*)—is arguably the ideal overall approach for qualitative studies, although frequently strategies are combined.

Convenience Sampling



Although convenience sampling is the easiest way to compile a sample, it is also the weakest one. Suppose you are interested in shoppers' opinions about rising prices as a result of inflation. If you stand outside a shopping center and hand out questionnaires as shoppers walk in, the sample of respondents you will obtain will be a convenience sample. This is a biased sample because you are systematically excluding people who shop at other places. The opinions that you would get from your chosen sample may reflect the distinct characteristics of this shopping center, such as the nature of its stores (high-end stores will attract a more affluent demographic) or its location (a shopping center close to a university will attract primarily university students with distinct purchasing habits). As a result, your sample will likely not be very representative of the opinions of the shopper population at large, and the scientific generalizability of your observations will be limited. *Melanie Pongratz, via Unsplash*

When researchers use **convenience sampling**, they draw their samples from a part of the population of interest that is close at hand or otherwise readily available to them. Often, sampling units are selected because they happen to be in the right place at the right time. Examples of convenience sampling would be sampling a population of workers by interviewing employees walking into an office building or sampling a population of patients by surveying those who stop by a medical clinic. For their part, student researchers are often enamored with obtaining convenience samples of students at their own university, which they can easily collect by hanging out on the quad or in the library and buttonholing people for interviews. You can see how this approach would lead to a skewed sample. First of all, are the students at your particular university a good stand-in for all college students? Certainly not—among other things, one university would reflect the views of students in just one part of the country. Even if you were interested only in the views of students at your one university, you'd be smart not to gather your sample on the quad or in the library—that approach would not generate a representative sample of all students, because certain types of students are more likely to congregate in those spaces.

Convenience samples are generally frowned upon by social scientists. (Another name for this technique is *haphazard sampling*, which better captures this negative impression.) Convenience sampling can be useful for pilot testing and in exploratory research, but generally speaking, you don't want to sample in such an arbitrary way. After all, this is basically how journalists gather sources for their stories. They need quick and easy access to people to talk about a particular topic (think of the staple television news interview with random people on the street), and they are less concerned with how well the opinions of their interviewees reflect those of the larger population of interest. Sociologists try to be more systematic in their approach. Although convenience samples offer one major benefit—convenience—we usually want to use a more rigorous strategy, and we should be cautious about generalizing from any research that relies on such samples.

Snowball Sampling

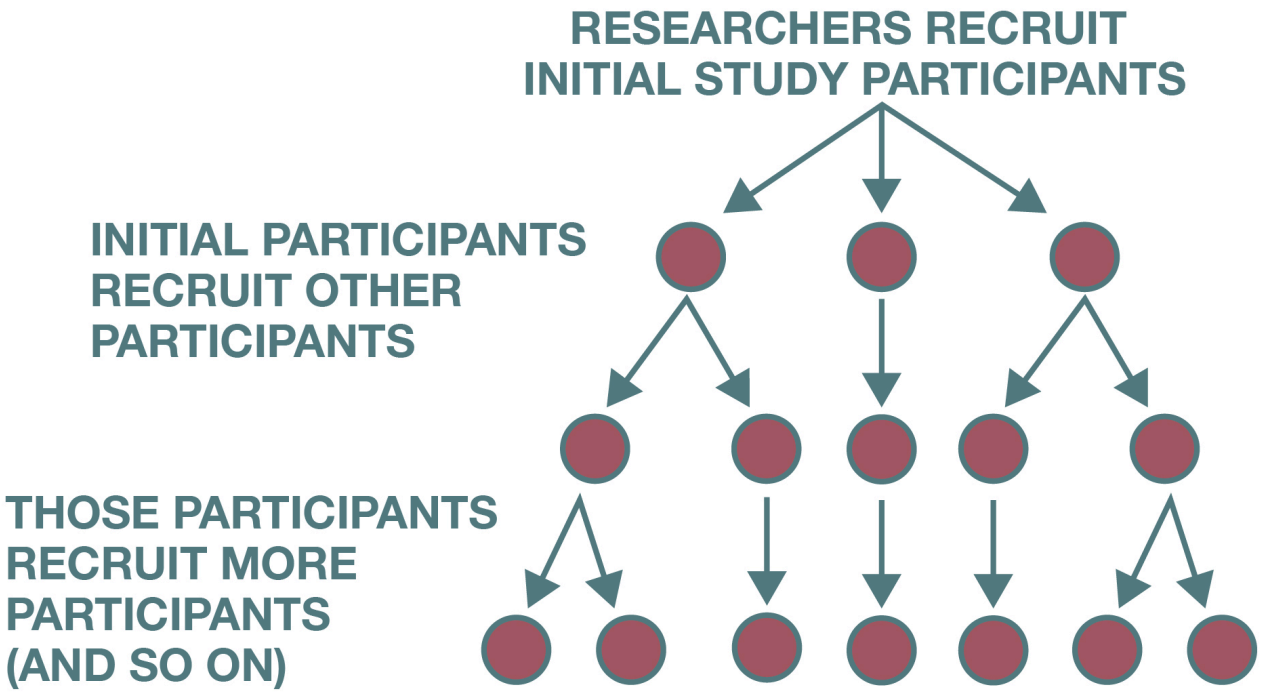


Figure 6.8. Snowball Sampling. In snowball sampling, you start by identifying a few respondents that match the criteria for inclusion in your study, and then you ask them to recommend others they know who also meet your selection criteria.

Researchers sometimes rely on **snowball sampling** techniques to identify study participants (see **Figure 6.8**). In this case, a researcher might already know of one or two people they want to include in their study but then might rely on those initial participants to help identify additional study participants. Based on these referrals, the researcher's sample builds and becomes larger as the study continues—much as a snowball builds and becomes larger as it rolls through the snow.

Snowball sampling is an especially useful strategy when a researcher wishes to study a hidden or hard-to-reach population or some stigmatized group or behavior. For example, a researcher who wants to study how undocumented immigrants access healthcare services would be unlikely to find many participants by posting a call for interviewees on a website or making an announcement about the study at some large social gathering. Instead, the researcher might know someone who is undocumented, interview that person, and then be referred by the first interviewee to another potential subject. Having a previous participant vouch for the trustworthiness of the researcher may help new potential participants feel comfortable about being included in the study.

Snowball sampling is sometimes referred to as *chain referral* sampling. One research participant refers another, and that person refers another, and that person refers another, and through this process, a chain of potential participants is identified. In addition to using this sampling strategy for potentially stigmatized populations, it is also a useful strategy to use when the researcher's group of interest is likely to be difficult to find—not only because of some stigma associated with the group but also because the group may be relatively rare. This was the case for Steven M. Kogan and his colleagues (2011). Kogan's research team wished to study the sexual behaviors of non-college-bound African American young adults who lived in high-poverty rural areas. The researchers first relied on their networks to identify study participants, but because members of the study's target population were not easy to find, access to the networks of initial study participants was important for identifying additional participants. Initial participants were given coupons to pass on to others they knew who qualified for the study. Participants were given an added incentive for referring eligible study participants; they received not only $50 for participating in the study but also $20 for each person they recruited who also participated in the study. Using this strategy, Kogan and colleagues succeeded in recruiting 292 study participants.

The problem with snowball sampling is the sampling bias it inevitably introduces because it relies on social networks and therefore gives people in those networks a greater chance of being selected for the sample (what is called *in-network selection*). This is a problem for a number of reasons. For one thing, because people tend to know and befriend other people like them, the participants you recruit through snowball sampling will likely be similar to one another, which means you may miss certain kinds of people in your population. Furthermore, using referrals will tend to exclude more socially isolated individuals from your sample, given that these potential respondents will not have as many people identifying them to you.

The sampling bias inherent to snowball sampling discourages its use among quantitative researchers in particular, who typically want more representative samples. That said, this approach is often used to augment a sample in qualitative research, and more technically advanced methods of snowball sampling have been developed to reduce the degree of sampling bias it creates.

Quota Sampling



Figure 6.9. Example of Quota Sampling. In quota sampling, units are selected into a sample on the basis of prespecified characteristics. Usually the goal is to have representation of the different groups that make up the population. However, the size of these groups in the sample does not usually represent the actual size in the larger population. In this example, the researchers want to ensure that the sample includes men, women, white individuals, and African American individuals. Because this is a nonprobability sampling strategy, it is important to capture the experiences of these specific groups, and not necessarily to have an accurate representation of the gender and racial distributions in the larger population (which likely includes nonbinary individuals and people of other races).

Quota sampling is a nonprobability sampling strategy that quantitative researchers use frequently (see **Figure 6.9**). Qualitative researchers also employ this strategy, although they more typically characterize what they do as *purposive* sampling or *theoretical* sampling (described in the next section)—an approach that integrates quota sampling as part of a larger strategy driven by a researcher's theories about the factors and processes that matter.

When conducting quota sampling, you first identify categories that are important to the study and for which some variation is likely. You create subgroups based on each category and decide on how many people (or other units of analysis) to include from each subgroup. You then collect data from that specified number of cases for each subgroup.

Suppose you are doing a survey of student satisfaction with on-campus housing. The two types of housing on your campus are apartments that include full kitchens, and dorm rooms in which residents do not cook for themselves but eat in a cafeteria. As a researcher, you might wish to understand how satisfaction varies across these two types of housing arrangements. Perhaps you have the time and resources to survey 100 campus residents, so you decide to survey 50 from each housing type. Let's say that your review of the literature on university housing also suggests that

campus housing experiences vary by gender. To address this potentially important factor in your study, you settle on four subgroups for your sample: men who live in apartments, women who live in apartments, men who live in dorm rooms, and women who live in dorm rooms. Your quota sample includes 25 people from each subgroup.

Note that the quota sample in this example is not a representative sample in the full sense of that term. It's likely not the case that the residents of on-campus housing at your university are split evenly between men and women and those in apartments and those in dorm rooms. So any statistics you calculate for your sample will not necessarily reflect those of the actual population of students living in on-campus housing. For this reason, researchers would normally prefer a probability sample, which would mean randomly sampling respondents from the entire population of students who live in on-campus housing. This process could get you sufficient numbers of people in each of the four subgroups just described through random selection alone, and you would not have to dictate a threshold of 25 respondents for each category.

That said, researchers may have reason to use quota sampling rather than probability sampling techniques. It may be difficult to recruit a truly random sample from the target population, and yet the researchers still will want to make sure they have adequate coverage of important kinds of variation in the sample. In this case, a nonrepresentative sample can be justified by theory—specifically, by theories of what those important factors are. In the campus housing example, you could justify your choice of four subgroups varying by gender and housing type based on your knowledge (from the literature or your preliminary observations) of what factors shape the experience of on-campus housing in noteworthy ways.

Furthermore, quantitative researchers can conduct quota sampling in ways that give them some confidence that their results are generalizable to the population of interest. For instance, political pollsters can construct a representative sample of voters in an election by recruiting sufficient numbers of people in important subgroups. Specifically, their samples might have quotas for Democrats and Republican voters, college-educated and non-college-educated voters, and white and nonwhite voters, given how much those factors are known to affect voting decisions. Each subgroup would be represented in the sample in proportions that match their proportions in the overall population—whose characteristics the pollster can uncover from census data or other surveys. If the pollster (1) is right about what characteristics really do shape voting decisions *and* (2) has accurate and up-to-date knowledge about the prevalence of those characteristics in the overall population, then their findings should be generalizable to that population: a poll based on this constructed sample will reflect voters' overall preferences.

Those two conditions are not always easy to achieve. For instance, George Gallup's coup in predicting the 1936 presidential election (discussed in the sidebar **Deeper Dive: Sampling Gone Wrong**) ensured that the approach he pioneered, quota sampling, dominated the field for a decade. But in 1948, Gallup and most other pollsters were wildly wrong about the outcome of that year's presidential election, as described in the next sidebar **Deeper Dive: Sampling Gone Wrong (Again)**. The polling industry subsequently turned to probability sampling techniques that were just being developed, no longer confident that they could regularly achieve the necessary conditions.

In more recent decades, however, quota sampling has experienced a resurgence of sorts. Rigorous phone surveys have become much more difficult and expensive to conduct (as we will discuss in **Chapter 13: Surveys**). As a result, pollsters are increasingly turning to *online* polls that use quota samples as an alternative to random samples gathered through phone polling. Modern-day quota sampling, however, uses a more sophisticated approach than what Gallup and his contemporaries had available to them. Rather than just dictating the recruitment of a certain number of cases in each relevant category, researchers also perform statistical weighting to adjust the sample after it has been collected—what is called **postsurvey weighting**. This weighting gives cases from underrepresented subgroups a greater impact on the results, which means that the statistics generated better reflect what one would find with a truly representative sample.

For example, the Pew Research Center conducts regular online surveys using its online panel, the American Trends Panel (ATP). They recruit their respondents offline by taking a random national sample of residential addresses—a

sampling frame, as we mentioned earlier, often used to generate representative samples. Once respondents agree to participate in the ATP, Pew gives them online surveys to take. It then conducts postsurvey weighting, with each survey's results "statistically adjusted to match national estimates for political party identification and registered voter status in addition to demographics and other benchmarks" (Kennedy et al. 2021). This **weighted sample** is meant to be representative of the national population.

Deeper Dive: Sampling Gone Wrong (Again)



Harry S. Truman holding the *Chicago Daily Tribune* with the erroneous headline, "Dewey Defeats Truman," at Union Station in St. Louis, Missouri, on November 4, 1948, after winning that year's U.S. presidential election. Truman was so roundly expected to lose that the Tribune printed the erroneous headline, boldly anticipating victory for his opponent, Thomas E. Dewey. Their false story was widely distributed by the Associated Press and published in numerous newspapers on November 5. *Byron H. Rollins, via Wikimedia Commons*

As we described in **Deeper Dive: Sampling Gone Wrong**, the magazine *Literary Digest* erroneously predicted in 1936 that Franklin D. Roosevelt, the Democratic incumbent running for reelection, would lose in a landslide. *Literary Digest* had recruited its survey respondents in a way that generated a huge—but woefully biased—sample. Survey techniques were rudimentary at this time, and pollsters did not yet have good strategies for obtaining a probability sample of U.S. voters. Nevertheless, up-and-coming pollster George Gallup successfully predicted the election's outcome. How did he do that? Using quota sampling, Gallup had figured out which characteristics—such as gender, age, household income, and area of residence—would be decisive in determining how Americans voted. He created subgroups based on these categories and set quotas for each one. Then he recruited respondents to fill his quotas, winding up with a sample that looked like the population—at least in terms of the specific characteristics he had identified as being decisive.

When Gallup's prediction in 1936 turned out to be correct, he instantly became the face of the new science of political polling. His polls started regularly appearing on the front pages of newspapers like the *Washington Post*. And his legend only grew after Gallup successfully called several subsequent presidential elections based on quota samples. Other pollsters, such as Elmo Roper and Archibald Crossley, used quota sampling with similar success.

In 1948, quota sampling's long streak of electoral success ended. That year, prominent polls based on quota samples—Gallup's included—incorrectly predicted that the Democratic incumbent Harry Truman would lose to

his Republican challenger Thomas Dewey.[1] Gallup was so sure about Truman's loss that his organization stopped polling three weeks before the election, even though 14 percent of voters indicated in polls that they were undecided. With many national polls predicting a Dewey victory, newspapers prepared editions of their papers for printing on the morning after the election with headlines declaring Dewey the victor and speculating on choices for his presidential cabinet. Truman ended up winning by 5 percentage points, and the morning editions were embarrassingly wrong—as captured in that iconic photo of the president-elect holding up a newspaper with the headline "Dewey Defeats Truman."

Various factors contributed to the fact that the polls were so off in 1948. It is possible that voter opinion had swung in the last weeks of the campaign. The polls may have captured the overall sentiment of the country's voters, but they failed to predict accurately who actually turned out to vote (a topic to be discussed further in **Chapter 13: Surveys**) (Lester 1998).

After that year's stunning miscall, pollsters made sure to survey voters up until election day and to do a better job identifying voters likely to get out and vote. But arguably the most dramatic change in polling after 1948 was the shift toward probability samples whose respondents were selected randomly from the population. Pollsters increasingly saw nonprobability sampling approaches as too risky. The chief weakness of quota sampling is that it relies on our *existing knowledge* of what the population of interest actually looks like across the characteristics *we believe* matter. When we don't have an accurate picture of the population and its decisive characteristics, our results can turn out to be wildly wrong (Neuman 2006). After 1948, pollsters worried that their quota categories were not adequately representing the electorate, and they gradually gave up on this approach. "Political polling was non-probability, and for a number of years they got away with it. In 1948, they got burned," pollster Warren Mitofsky told the Associated Press in a retrospective on that critical presidential election (Lester 1998). In the 1960s, Mitofsky would pioneer the use of phone surveys using random-digit dialing, the polling method that eventually succeeded quota sampling.

Purposive Sampling

By now you might have a justifiably dim view of nonprobability sampling. We noted the serious problems of representativeness that plague samples drawn from convenience sampling, snowball sampling, and quota sampling. Why would any sociologist in their right mind use a nonprobability sample? Simply put, qualitative researchers do so because their methods prioritize depth in their research over breadth—in-depth knowledge rather than generalizability. In probability sampling, the selection of cases is random, given the goal of generating representative samples whose statistics are generalizable to the population of interest. In *purposive* sampling (sometimes referred to as *theoretical* sampling), the selection of cases is guided by some larger purpose on the part of the researcher—specifically, their theory about what concepts and processes matter.

Typically in qualitative studies, purposive sampling across at least two levels of analysis is required. First, the researcher needs to specify the **field site** they wish to examine—the location or grouping chosen for study to understand a phenomenon present within a larger context. For the purposes of sampling, we are using the term *site* in a broad sense to refer to the specific locations or groupings from which our micro-level data (e.g., in-depth interviews with individuals situated in those locations or groupings) will come. This sort of staged approach is necessary because unlike quantitative

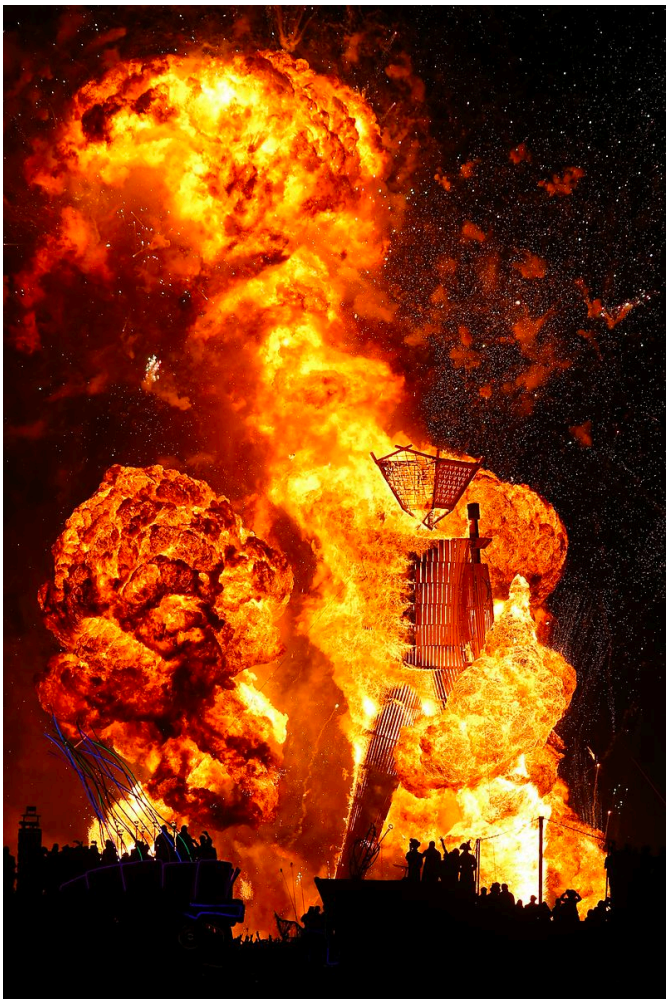
1. The PBS documentary *The First Measured Century* covers the 1948 election and other historically significant events related to social scientific measurement.

researchers, qualitative sociologists would not study a random sample of Americans (or Virginians) to understand what Americans think. Instead, they would choose a city or community to study—maybe more than one, if they are doing a larger study—and then find people to interview or observe in that locality. Rather than studying a random national sample of organizations, they would choose one or more specific groups (perhaps located in one or more specific localities, too). Rather than studying a random national sample of workers, they would choose one or more specific occupations or companies or workplaces. The hope is that the specific localities, organizations, or occupations they pick would be able to tell them something about Americans, organizations, or workers more broadly.

In purposive sampling, the site is always a stand-in, or representative, for something larger. (Indeed, qualitative researchers often use the term *case selection* when they are talking about their choice of a particular site to study: they want to argue on behalf of that specific locality, organization, or occupation as an ideal or useful case of that larger phenomenon.) For instance, if we are studying evictions in Milwaukee, we aren't just interested in that city's eviction problem—ultimately, we want to say something about evictions in all cities, as the sociologist Matthew Desmond (2016) was able to do in his Pulitzer Prize–winning ethnography on the subject. That said, it is important to emphasize that the site a qualitative researcher chooses will never be "representative" in the sense we have been using throughout this chapter. That would require a much larger sample and some kind of probability sampling approach.

With any luck, however, the site will be useful and informative in generating and refining theories. We will have more to say about site selection in **Chapter 9: Ethnography**, but the most important thing to note is that some purpose should justify your site selection when you are using a theoretical sampling approach. Answer the question, "What is this a case of?" Specifically, you want to argue on behalf of what makes your choice of a site interesting and potentially productive. (Don't *just* say you did it because the locality happened to be where you live, or the organization happened to be one you volunteer for, even though feasibility and access are obviously important reasons for choosing a particular site, too.) At the same time, you don't have to choose a site that is "average" across all the relevant variables—say, the "average" U.S. city in terms of housing prices, or the "average" company in terms of workforce size and revenue, or the "average" occupation in terms of wages. If you are studying one site or even several sites, there is no way your case selection will be truly representative of all of the possible localities, organizations, or occupations to study—there is just too much variation to account for. Future qualitative research can tackle other sites, and future quantitative research can give us a bird's-eye perspective with a representative sample. Your goal is to produce research that says something meaningful about this site and can plausibly be true across other sites as well.

Granted, some qualitative sociologists have made a name for themselves by studying what they argued was the quintessentially "typical" instance of a site. Robert S. Lynd and Helen Merrell Lynd (1929, 1937) conducted an influential research project on Muncie, Indiana—which they called "Middletown" in their books—to understand how American culture was adapting to dramatic economic and technological change. That study was replicated in the late 1970s and again in 1999 (Caplow 1982, 1983), and as a result of this research, Muncie has gained a reputation for being the "typical" American small city. That said, rather than going on a dubious quest for a somehow "ideal" or "exemplary" site, it is often more theoretically useful to examine what Katherine K. Chen (2015) has called **extreme cases**. These are examples of sites that display characteristics or behaviors that are *not* the norm—they are **outliers**, or **deviant cases**, that stray from commonly seen patterns in instructive ways. (While Chen used the term "extreme cases" to apply to organizations, we could apply this logic to any kind of site.) As Chen points out, extreme organizations can show us processes or perspectives that we rarely see in more typical groups, making it easier to study those phenomena. Because of their extraordinary nature, making them the focus of study can highlight practices that are taken for granted in most organizations. And they can present intriguing puzzles or interesting tensions because of their uniqueness. For example, Chen (2009, 2012) conducted a study of the Burning Man organization, which puts on an annual arts festival in Nevada. Burning Man is not your typical organization—or even your typical arts festival—but that fact did not make Chen's study a pointless academic exercise. On the contrary, Burning Man provided a fascinating case of an organization that had managed to rapidly grow even though it rejected the conventional approach that most organizations took to survive and succeed: formal bureaucratic rules, hierarchical structure, and centralized power. Burning Man was not ordinary, and its exceptional nature needed to be explained—which, in turn, provided insight into possibilities for democratic organizing and consensus-based decision-making in organizations more broadly.



Using the Burning Man organization as an *extreme case*, sociologist Katherine K. Chen investigated how this organization managed to grow rapidly despite lacking the bureaucratic elements of conventional organizations. *torroid, via Flickr*

Particularly ambitious researchers may decide to compare and contrast two (or more) sites. Investigating multiple sites allows the researcher to see, among other things, how changing a key set of variables may lead to different outcomes. For instance, Tomás Jiménez (2008) sought to examine the role that "immigrant replenishment"—the continual migration of immigrants from a particular country—played in sharpening the ethnic identities of third-generation Mexican Americans. Jiménez's purposive site sampling operated on multiple levels. He chose immigration from Mexico—out of all the possible countries to study—because the experience of later generations of Mexican Americans differed sharply from that of the descendants of European immigrants in the twentieth century, given that European immigration dwindled over that time. In other words, Mexican immigration was Jiménez's extreme case—an outlier

that challenged prevailing understandings of the expected progress toward assimilation and incorporation among the descendants of immigrants. In designing his interview-based study, however, Jiménez also had to make site decisions regarding the places where he would conduct his interviews. He chose two cities—Garden City, Kansas, and Santa Maria, California—which both were currently experiencing replenishment of their Mexican migrant populations. However, the replenishment in Garden City had only recently begun again after a long period of little immigration, whereas in Santa Maria, the replenishment had been continuous for generations. As it turned out, the experiences of Jiménez's interviewees were the same across these two sites. This often happens in research—what you thought might matter doesn't turn out to matter—and this null finding can itself be interesting. In any case, Jiménez's decision to study two cities that were both experiencing immigrant replenishment had a clear purpose—namely, to see if varying the *kind* of immigrant replenishment would matter.



In a seminal sociological study, Arlie Hochschild used the cases of flight attendants and bill collectors to introduce the concept of *emotional labor*. The selected professions reveal how workers in the service industry adjust their emotions to further the goals of their employers. *Vinh Lâm, via Pexels*

A classic example of a comparative-site study is Arlie Hochschild's book *The Managed Heart* (2012), which examined how workers engaged in *emotional labor*, the effort put into modulating their expressed emotions on the job to provide superior customer service or further other goals of their employers. For this qualitative study, Hochschild chose to observe and interview two types of workers (two "sites" in the broad sense we are using): flight attendants and bill collectors. Her case selection was particularly ingenious. By looking at these two professions, she could see either side of the spectrum of emotional labor: pleasing customers and being incessantly cheerful and helpful (flight attendants)

versus berating and intimidating debtors to coerce them to pay up (bill collectors). Here, too, theory drove the selection of the sites. A similar but more recent ethnography by Benjamin Snyder (2016) looks at three groups—long-haul truck drivers, financial professionals, and unemployed jobseekers—to understand how workers experience time in their day-to-day lives. Again, these three sites are not "typical" of all workers, but each category highlights the very different ways that individuals must adapt their bodies and minds to the rhythms of the modern economy.

After qualitative researchers decide on sites to study, they then need to apply theoretical sampling techniques to their choice of individuals to interview or places and events to observe. One useful strategy is to **sample for range**. In this strategy, the sample selection is driven by a theory of what variables matter (hence, theoretical sampling). The researcher tries to find enough cases with variation across those variables. Oftentimes, qualitative sociologists will expressly identify certain subgroups that they want to sample that span the desired range of variation.

In practice, a qualitative sociologist's strategy of sampling for range looks like a mash-up of stratified sampling and quota sampling. Previously, we described stratified sampling as a probability sampling method, but the same basic principles can be applied to nonprobability samples—that is, dividing the population into subgroups, or strata, and then finding cases that fit those subgroups. As with quota sampling, the researcher might have a certain number of cases in mind for each of those subgroups. For example, for her study of unemployment, Sarah Damaske (2020, 2021) conducted 100 in-depth interviews with unemployed individuals in Pennsylvania. She employed a "stratified theoretical sampling strategy" (2020:8) that drew interviewees from four groups: 26 middle-class men, 18 middle-class women, 25 working-class men, and 31 working-class women. ("Class" was defined by whether or not interviewees had gone to college.) Note that this strategy of sampling within subgroups, or strata, could involve looking for "extreme" categories—say, sampling flight attendants and bill collectors to represent either end of the emotional labor spectrum of friendly and hostile work personas. But sampling for range does not have to cover such a broad sweep of variation, as Damaske's research shows.

Indeed, purposive sampling may be an effective approach when a researcher wishes to include only people who meet narrow or specific criteria. For example, Amy Non and her colleagues (2019) were interested in learning about the stress experiences of Mexican women who are mothers and live in a new immigrant destination in the United States. They conducted a series of focus groups with women who fit their **inclusion criteria**: a self-described Mexican-born immigrant, age 18 or older, who had at least one child and lived in Nashville, Tennessee. In this case, the researchers' goal was to find participants with specific experiences rather than trying to cover the entire spectrum of experiences.

When conducting purposive sampling, you need to make numerous decisions about both the inclusion criteria and **exclusion criteria** you will use—that is, how you will decide whether to include or exclude a person (or other unit of analysis) from your sample. Sometimes, you may decide to focus on just a particular subgroup of a larger population because that is the group you find most theoretically interesting or you wish to reduce the amount of variation so that other factors can be better observed. For instance, in his research on the reasons that working-class individuals are leaving the labor force, Robert Francis (2021) focused on working-class men rather than women, given how that trend has been more pronounced in recent years among men with less education. The exclusion criteria for his sample of in-depth interviews, therefore, explicitly excluded women. In Victor's study of long-term unemployment (V. T. Chen 2015), he interviewed and observed unemployed workers who previously worked at auto plants in Detroit and Windsor, Canada—two cities across the river from one another. The inclusion and exclusion criteria were intended to provide more of an apples-to-apples comparison between the two groups being compared, who would be largely the same except that they lived on either side of the border, which would allow Victor to focus on how the different social policies in the United States and Canada made a difference in these workers' lives. While narrowing the sample in these ways can be useful, we would not suggest getting too hung up over exclusion and inclusion criteria in purposive sampling—an advantage of this approach is that you can find *unexpected* cases that might not support your existing view of the phenomenon but are insightful nonetheless.

It is helpful to think of sampling as an evolving *process* in qualitative research. You do not sample all at once with a set plan in mind, as you would for a quantitative study. Instead, you should let theory guide you—from interviewee to

interviewee, from observation to observation. Each case will offer new patterns that you can explore in subsequent interviews or observations. When another case fits the pattern—when the same causal mechanisms seem to be operating—you have some confirmation of your theory. When another case deviates from the pattern, you can find a way to explain it, refining and expanding your original theory in the process. To illustrate this approach of "sequential interviewing" (which obviously can be applied to other qualitative methods as well), Small (2009) provides a hypothetical example of a researcher, Jane, who is studying African American attitudes toward immigration. One interviewee tells Jane about experiencing discrimination from Latino immigrants when she was younger, which has made that person favor stricter immigration laws. This prompts Jane to formulate a theory linking personal encounters with discrimination from immigrants with later support of immigration reform. Jane tests this theory in future interviews by seeing if it holds up—replicates—across different contexts. Specifically, she finds other African American interviewees with similar experiences and sees if the causal process identified earlier replicates in their cases—that is, if encounters with discrimination also make them more likely to support immigration reform. Jane also finds African American interviewees who have not had such experiences, and those who experienced discrimination from another group, such as Russian immigrants. She refines her original theory to accommodate what those interviewees tell her: whether the absence of discrimination leads to less support of immigration reform, whether discrimination from other groups also shapes those policy views, and so on.

This approach of replicate-as-you-go plays to the strengths of the qualitative method. Unlike in quantitative studies, you can refine your theoretical focus as you add more cases to your sample—that is, as you do more observations or interviews. In this process of **iterative sampling** (or *recursive sampling*), researchers move back and forth between the process of sampling and the preliminary analysis of data—with their analysis informing their selection of new cases for their sample, and each new case providing more data to be analyzed (Cohen and Crabtree 2006). You can seek out specific cases to test the new hunches you develop in the field or to capture more variation across the variables you learn are important. As you sample, you should continually ask the key question, "What is this a case of?" That will help you understand the significance of each interview you do or observation you make, possibly generating additional theories to test or explore.

If you happen to stumble across a deviant case that doesn't fit your current theory, you can seek out more data, including more cases, that might explain that deviation. Indeed, this search for *instructive outliers* is a key advantage of a purposive sampling strategy: you can find puzzling cases that surprise you, ones that you did not anticipate when you began your study, and the existence of each instructive outlier speaks to the need to refine existing theories to accommodate it.[2] (Sociologists call this approach of looking for surprising cases *abduction*, a topic to be discussed in **Chapter 11: Qualitative Data Analysis**.) Being able to move back and forth between sampling and analysis—the recursion involved in such an approach—allows qualitative researchers to focus on this important theoretical work.

As we discussed, a qualitative study typically continues this iterative process until the researcher reaches the point of saturation—when they are not gaining any more useful knowledge from each new case. In purposive sampling, this typically means that you have covered all the important forms of variation (perhaps as specified subgroups within your

2. Sometimes, researchers call these outliers "negative cases"—"negative" in the sense they do not follow the expected pattern; an example would be the previously mentioned immigration study by Tomás Jiménez that focused on the "negative case" of Mexican immigration (in which assimilation patterns seen in previous European immigration do not hold, arguably because of immigration replenishment in the Mexican case). When outliers are particularly instructive, a single case can challenge existing understandings of a phenomenon. The term "black swans" is used to describe these situations (Taleb 2010). Europeans thought black swans were fictional creatures until they were actually spotted in Australia; that one sighting transformed existing thinking. In a similar way, an instructive outlier can shake and even overturn a prevailing theory, including any reasoning that followed from it.

sample). Qualitative interviewing guru Robert Weiss (2004) suggests that a qualitative researcher will need at least five instances of any important "types," or subgroups, within their sample. That means for a study with "between a half dozen and a dozen important types of respondents," which is a typical solo-authored qualitative project, a "rough ideal" would be 30 to 60 respondents in your sample (Weiss 2004:48).

In Table 6.2, we review the situations when you might want to use nonprobability sampling and the drawbacks associated with each strategy.

Table 6.2. When to Use Nonprobability Sampling Strategies

Sampling Strategy	Suggested Use	Key Problems
Convenience sampling	When you want to quickly generate a sample for pilot-testing questions or conducting a pilot study	Major problems with sample representativeness
Snowball sampling	When the nature of the topic or the population you are studying makes it difficult to recruit respondents or get them to talk candidly with you	People in the same social networks tend to be similar, meaning you may miss important variation Socially isolated individuals will be less likely to be in your sample
Quota sampling	When you have key subgroups of your population in mind and want to make sure you have sufficient numbers of cases in each	Requires weighting and/or sample selection in line with relevant population characteristics for the sample to be representative
Purposive sampling (Theoretical sampling)	When you want to follow the ideal approach for most qualitative studies When you want to supplement your strategy with other probability and nonprobability sampling techniques	If you really need a representative sample, then this is not the best approach (and you should avoid other nonprobability sampling strategies, too)

Key Takeaways

1. Nonprobability samples might be used when researchers are conducting exploratory research, pilot studies, or qualitative research. They are especially useful when the researcher's focus is theoretical and their goal is to obtain an in-depth understanding rather than a more generalizable understanding of an issue.
2. The various types of nonprobability samples include purposive samples, snowball samples, quota samples, and convenience samples.
3. A researcher's findings need not be generalizable to be valuable; samples that allow for comparisons of theoretically important concepts or variables may yield findings that contribute to our social theories and

our understandings of social processes.

Exercises

1. Of the four nonprobability sample types described, which seems strongest to you? Which seems weakest? Explain.
2. Imagine you are about to conduct a study of people's use of the public parks in your hometown. Explain how you could employ each of the nonprobability sampling techniques described previously to recruit a sample for your study. Compare this approach to the random sampling techniques you used for **Exercise 2** in **Section 6.3**.
3. Now let's say you wanted to do a qualitative interview-based study of U.S. college students to understand the effect of drug use on academic performance. How would you obtain a purposive sample? Compare this approach to the techniques you proposed for **Exercise 2** and for the probability sampling strategy you suggested for the similarly themed study of drug use for **Exercise 3** in **Section 6.3**.

6.5. Questions to Ask about Samples

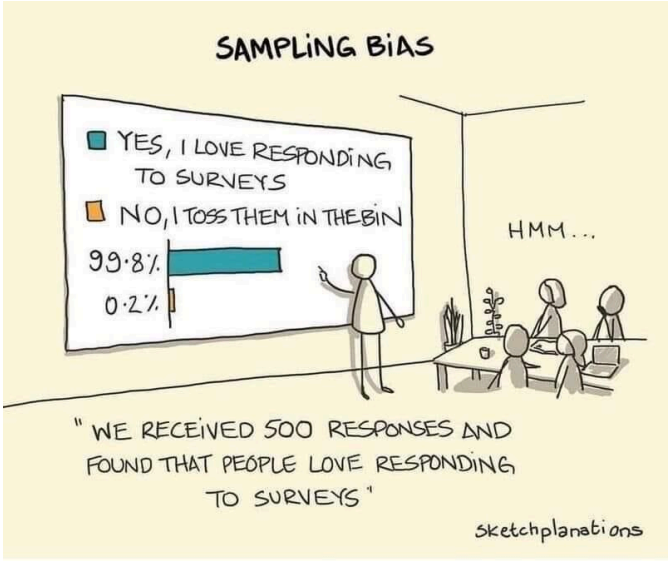
Learning Objectives

1. Identify several questions we should ask about samples when reading the results of research.
2. Name some tenets worth keeping in mind with respect to responsibly reading research findings.

We read and hear about research results so often that we might overlook the need to ask important questions about where research participants come from and how they are identified for inclusion in a research project. It is easy to focus only on findings when we are busy and when the really interesting stuff is in a study's conclusions, and not its procedures. But now that you have some familiarity with the variety of procedures used to select study participants, you are equipped to ask critical questions about the findings you read.

The first thing to ask is how the sampling was conducted. We have talked about the various probability and nonprobability sampling techniques and their strengths and weaknesses. With the study's research question and overall approach in mind, you should have a sense of whether the researchers want to use their sample to make claims about the prevalence of certain characteristics or activities in the target population. If so, then a probability sample (or a quota sample devised to be representative) is called for. A study that doesn't use a probability sampling technique is unlikely to give us results that we can generalize to the population of interest with much confidence. A probability sampling approach, however, probably doesn't make much sense for a qualitative study that is focusing on developing complex theories and generating a more granular understanding of a phenomenon. With their inherently small samples, qualitative studies are better off pursuing a purposive sampling approach grounded in theories about what factors matter and allowing for in-depth investigation of causal mechanisms.

Sketch cartoon of a person pointing to the results of a survey in which 99.8 percent of respondents indicated that they love responding to surveys.

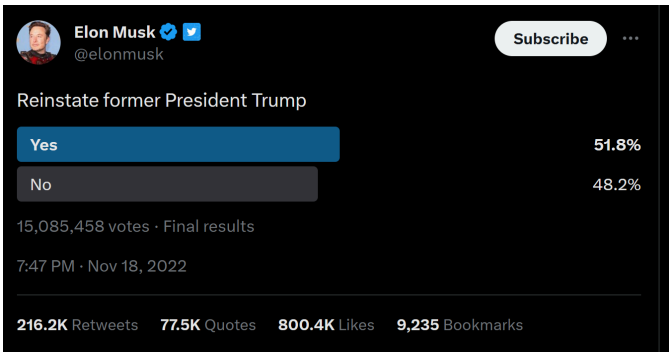


Sampling biases make the sample unrepresentative of the intended population and hurt generalizability claims about inferences drawn from the biased sample. In this example, if you hope to receive responses to your research questions through surveys, then you will get responses from people who happen to not mind doing surveys. From anyone who throws them in the bin, you will not receive a single data point. *Jono Hey, via* Sketchplanations

As a critical reader of research, you should also be aware of the many forms of sampling bias and how they are rampant in polls—scientific and unscientific—that are publicized widely. We already discussed the various ways that flawed samples can fail to faithfully capture what is going on in the target population. Every day, a news website will publish another survey of people's opinions and self-reported behaviors. But how was the sample collected? Was it truly a random sample of the population of interest? Were some sorts of people more likely to respond than others? These questions should be asked of every sample.

Online polling is particularly fraught with these problems of bias (notwithstanding the quality online surveys we mentioned earlier that use quota sampling). As much as news and entertainment sites love hosting their own polls, virtually every one you take is fundamentally flawed from the perspective of scientific sampling. Let's say a local site is asking for the public's opinion about some community issue. Although the target population is the local community, the sampling frame would be much narrower—basically, every person who visited that site's page with the poll. Particular groups of people in the local community would be underrepresented in the sample that is drawn from this sampling frame: those without access to the internet, those who do not read that news website, and those who dislike answering online polls.

Another issue with a single-issue online poll like this one is that people with strong opinions on the issue are more likely to participate. This is an example of what we call **self-selection bias**—that certain types of people are more likely to volunteer for (select into) the sample. Regardless of how many thousands of people end up taking this poll, then, it would be a dubious measure of the public's actual views—much like *Literary Digest*'s poll of *millions* of U.S. voters turned out to be way off because of poor sampling.



The typical online poll—like this one that Twitter owner Elon Musk posted about whether to reinstate the banned account of former president Donald Trump—provides useless information because it will never give us a representative sample. Among other things, people with strong opinions on the subject are more likely to participate in the poll, which means the poll's results will be biased. X

Self-selection bias is similar to a form of sampling bias we mentioned earlier, nonresponse bias. This bias occurs when the people who chose not to answer our survey (nonresponders) look different from those who did respond (respondents) in terms of the variables that matter in our study. Even careful random sampling cannot avoid this problem. For example, in a mail survey on alcohol consumption, researcher Vivienne Lahaut and colleagues (2002) found that only about half the sample responded after the initial contact and two follow-up reminders. The danger was that the half who responded might have different patterns of alcohol consumption than the half who did not, which could lead to inaccurate conclusions on the part of the researchers. In other words, it was not only that the responders were potentially different from nonrespondents but also that they differed across one of the study's key variables—alcohol consumption.

To test for nonresponse bias, the researchers later made unannounced visits to the homes of a subset of the nonresponders—coming back up to five times if they did not find them at home. They found that the original nonresponders included an especially high proportion of abstainers (nondrinkers), which meant that their estimates of alcohol consumption based only on the original responders were too high. (We'll discuss methods for addressing nonresponse bias in **Chapter 13: Surveys**.)

Even when reputable polling organizations like the Pew Research Center conduct rigorous sampling, potential for bias continually creeps in. For example, a data quality analysis that Pew conducted on its online panel survey found that new participants were less Republican than those in prior waves, and individuals who lived in politically balanced areas were more likely to join than those in the country's most and least pro-Trump areas (Kennedy et al. 2021). Furthermore, Trump voters were more likely to drop out of the panel. (The latter problem is called **attrition bias**: those who leave the study come from particular subgroups, thereby undermining the sample's overall representativeness.) Once it identified these biases, Pew implemented strategies to modify recruitment and improve data quality. Not all pollsters do that.

With all the potential sources of sampling bias, how do we know when we can count on the results that are being reported to us? We can keep a couple of things in mind as we read the claims researchers make about their findings. First, remember that sample quality is determined only by the sample actually obtained, not by the sampling method. We may set out to administer a survey to a representative sample by correctly employing a random selection technique, but if only a handful of the people sampled actually respond to the survey (potential nonresponse bias), we will have to be very careful about the claims we make about our survey findings. As we mentioned earlier, researchers may also give into the temptation to talk about implications of their findings as though they apply to some group other than the population actually sampled. Though this tendency is usually quite innocent and does not come from a place of malice, knowledgeable consumers of research findings need to be attentive to this sort of bait and switch.

Conversely, you should be careful about critiquing studies—especially qualitative studies—by just pointing out their samples are not representative and stopping there. As we have repeated throughout this chapter, qualitative research in general does not generate representative samples. And even with quantitative studies, one could always point out that the same study could be done elsewhere—above all, in more non-WEIRD societies—to test whether the same findings hold up in that new context. Sample representativeness and study generalizability are critical concerns, but we want to judge a study based on what it is trying to do. Especially for qualitative studies, the more pertinent question is

whether we expect the findings to change if the study was conducted in a different locality or among a different set of individuals. In some cases, we will have many reasons to suspect that the study's results would not hold elsewhere—the site that was sampled is just too distinct. But oftentimes a good case can be made, even with qualitative projects with nonrepresentative samples, that the patterns and processes we observed in one context could plausibly be found in other contexts, too. For a theoretically sampled study, this claim might be enough—and we can always conduct more research to test out the theories we generated.

At their core, questions about sample quality should address who has been sampled, how they were sampled, and for what purpose they were sampled. Being able to answer these questions will help you better understand, and more responsibly read, research results.

Key Takeaway

Sometimes researchers may make claims about populations other than those from whom their samples were drawn; other times they may make claims about a population based on a sample that is not representative. As consumers of research, we should be attentive to both possibilities.

Exercise

Find a news story or blog entry that describes results from a social scientific study. How much detail is reported about the study's sample? What sorts of claims are made about the study's findings, and to whom do they apply?

6.6. References

Arnett, Jeffrey J. 2008. "The Neglected 95%: Why American Psychology Needs to Become Less American." *American Psychologist* 63(7):602–14. doi: 10.1037/0003-066X.63.7.602.

Begley, Sharon. 2010. "The Trouble with Using Undergrads for Research." *Newsweek*, July 23.

Brockell, Gillian. 2020. "The History of Pollsters Blowing It Decades before Trump vs. Biden." *Washington Post*, November 6.

Caplow, Theodore. 1982. *Middletown Families: Fifty Years of Change and Continuity*. Minneapolis: University of Minnesota Press.

Caplow, Theodore. 1983. *All Faithful People: Change and Continuity in Middletown's Religion*. Minneapolis: University of Minnesota Press.

Case, Anne, and Angus Deaton. 2020. *Deaths of Despair and the Future of Capitalism*. Princeton, NJ: Princeton University Press.

Chen, Katherine K. 2009. *Enabling Creative Chaos: The Organization behind the Burning Man Event*. Chicago: University of Chicago Press.

Chen, Katherine K. 2012. "Charismatizing the Routine: Storytelling for Meaning and Agency in the Burning Man Organization." *Qualitative Sociology* 35(3):311–34. doi: 10.1007/s11133-012-9229-1.

Chen, Katherine K. 2015. "Using Extreme Cases to Understand Organizations." Pp. 33–44 in *Handbook of Qualitative Organizational Research: Innovative Pathways and Methods*, edited by K. D. Elsbach and R. M. Kramer. New York: Routledge.

Chen, Victor Tan. 2015. *Cut Loose: Jobless and Hopeless in an Unfair Economy*. Oakland: University of California Press.

Cohen, Deborah, and Benjamin Crabtree. 2006. "Iterative Sampling." *Qualitative Research Guidelines Project*. Retrieved December 30, 2022 (http://www.qualres.org/HomeIter-3827.html).

Damaske, Sarah. 2020. "Job Loss and Attempts to Return to Work: Complicating Inequalities across Gender and Class." *Gender and Society* 34(1):7–30. doi: 10.1177/0891243219869381.

Damaske, Sarah. 2021. *The Tolls of Uncertainty: How Privilege and the Guilt Gap Shape Unemployment in America*. Princeton, NJ: Princeton University Press.

Desmond, Matthew. 2016. *Evicted: Poverty and Profit in the American City*. New York: Broadway Books.

Francis, Robert. 2021. "Missing Men? Precarity and Declining Labor Force Participation among Working-Class Men." *Journal of Working-Class Studies* 6(2):65–79. doi: 10.13001/jwcs.v6i2.6835.

Henrich, Joseph, Steven J. Heine, and Ara Norenzayan. 2010. "The Weirdest People in the World?" *Behavioral and Brain Sciences* 33(2–3):61–83. doi: 10.1017/S0140525X0999152X.

Hochschild, Arlie Russell. 2012. *The Managed Heart: Commercialization of Human Feeling*. 3rd ed. Berkeley: University of California Press.

Holt, Jessica L., and Wayne Gillespie. 2008. "Intergenerational Transmission of Violence, Threatened Egoism, and

Reciprocity: A Test of Multiple Psychosocial Factors Affecting Intimate Partner Violence." *American Journal of Criminal Justice* 33(2):252–66. doi: 10.1007/s12103-008-9036-0.

Jiménez, Tomás R. 2008. "Mexican Immigrant Replenishment and the Continuing Significance of Ethnicity and Race." *American Journal of Sociology* 113(6):1527–67. doi: 10.1086/587151.

Kennedy, Courtney, Jesse L. Lopez, Scott Keeter, Arnold Lau, Nick Hatley, and Nick Bertoni. 2021. *Confronting 2016 and 2020 Polling Limitations*. Washington, DC: Pew Research Center.

Kogan, Steven M., Cyprian Wejnert, Yi-fu Chen, Gene H. Brody, and LaTrina M. Slater. 2011. "Respondent-Driven Sampling with Hard-to-Reach Emerging Adults: An Introduction and Case Study with Rural African Americans." *Journal of Adolescent Research* 26(1):30–60. doi: 10.1177/0743558410384734.

Lahaut, Viviënne M. H. C. J., Harrie A. M. Jansen, Dike van de Mheen, and Henk F. L. Garretsen. 2002. "Non-response Bias in a Sample Survey on Alcohol Consumption." *Alcohol and Alcoholism* 37(3):256–60. doi: 10.1093/alcalc/37.3.256.

Lester, Will. 1998. "'Dewey Defeats Truman' Disaster Haunts Pollsters." *Los Angeles Times*, November 1.

Lusinchi, Dominic. 2012. "'President' Landon and the 1936 Literary Digest Poll: Were Automobile and Telephone Owners to Blame?" *Social Science History* 36(1):23–54. doi: 10.1017/S014555320001035X.

Lynd, Robert S., and Helen Merrell Lynd. 1929. *Middletown: A Study in Modern American Culture*. New York: Harcourt, Brace and Company.

Lynd, Robert S., and Helen Merrell Lynd. 1937. *Middletown in Transition: A Study in Cultural Conflicts*. New York: Harcourt, Brace and Company.

Neuman, W. Lawrence. 2006. *Basics of Social Research: Qualitative and Quantitative Approaches*. 2nd ed. Boston: Pearson.

Non, Amy L., Gabriela León-Pérez, Holly Glass, Emma Kelly, and Nanibaa' A. Garrison. 2019. "Stress across Generations: A Qualitative Study of Stress, Coping, and Caregiving among Mexican Immigrant Mothers." *Ethnicity and Health* 24(4):378–94. doi: 10.1080/13557858.2017.1346184.

Selvin, Hanan C. 1958. "Durkheim's Suicide and Problems of Empirical Research." *American Journal of Sociology* 63(6):607–19. doi: 10.1086/222356.

Small, Mario Luis. 2009. "'How Many Cases Do I Need?': On Science and the Logic of Case Selection in Field-Based Research." *Ethnography* 10(1):5–38. doi: 10.1177/1466138108099586.

Snyder, Benjamin H. 2016. *The Disrupted Workplace: Time and the Moral Order of Flexible Capitalism*. New York: Oxford University Press.

Sohoni, Deenesh, and Jennifer Bickham Mendez. 2014. "Defining Immigrant Newcomers in New Destinations: Symbolic Boundaries in Williamsburg, Virginia." *Ethnic and Racial Studies* 37(3):496–516. doi: 10.1080/01419870.2012.716521.

Squire, Peverill. 1988. "Why the 1936 Literary Digest Poll Failed." *Public Opinion Quarterly* 52(1):125–33. doi: 10.1086/269085.

Taleb, Nassim Nicholas. 2010. *The Black Swan: Second Edition: The Impact of the Highly Improbable Fragility*. New York: Random House.

Weiss, Robert S. 2004. "In Their Own Words: Making the Most of Qualitative Interviews." *Contexts* 3(4):44–51. doi: 10.1525/ctx.2004.3.4.44.

7. MEASURING THE SOCIAL WORLD



So far, you have learned how to generate a working research question to guide the design of your proposed study. You've also learned how to identify a sample to study to answer that question. What happens next?

The first step is the **conceptualization** process, which we touched on in **Chapter 3: The Role of Theory in Research**. As a review, let's walk through the process of conceptualization for a hypothetical research question. Imagine you were going to embark on a project to research students' perceptions of traditional male and female gender roles. How would you go about conceptualizing what you mean by "gender roles"? Here is one plan of action:

1. Consider what you really want to know about the topic of male-female gender roles. You could start by brainstorming ideas with other students and reading some literature reviews on the topic (see **Chapter 5: Research Design**).
2. Your efforts should lead to the gradual development and refinement of a research question. As mentioned in **Chapter 4: Research Questions**, this process is *iterative*—that is, you will continue to edit and improve your research question, particularly as you consult the relevant literature. Let's say that the research question you eventually come up with is, "What is the relationship between students' attitudes about traditional gender roles and their future career plans?"
3. As you refine your research question, you will want to identify any concepts that relate to it, along with any relevant **dimensions** of those concepts (see the discussion of multidimensional concepts in **Chapter 3: The Role of Theory in Research**). You should then define your key concepts of interest. For the research question we just proposed, you might define "traditional gender roles" as the "commonly expected characteristics and behaviors determined by an individual's socially assigned gender at birth." "Career plans" might be defined as the "profession that a person ultimately hopes to pursue."
4. Create a conceptual map (see **Chapter 3: The Role of Theory in Research**) of how the concepts you have described might relate to one another. Ideally, you will be able to use existing theories and research results to get an idea of the possible relationships at hand.

After you clarify the nature of the concepts relevant to your research question, what do you do next? **Figure 7.1** shows

you the process that typically unfolds for researchers. After conceptualization, comes **operationalization**. (Yes, we agree, this word is a mouthful—like "Mississippi," it takes some practice for it to roll off your tongue. For starters, put your stress on the *ra* and *za* syllables.)

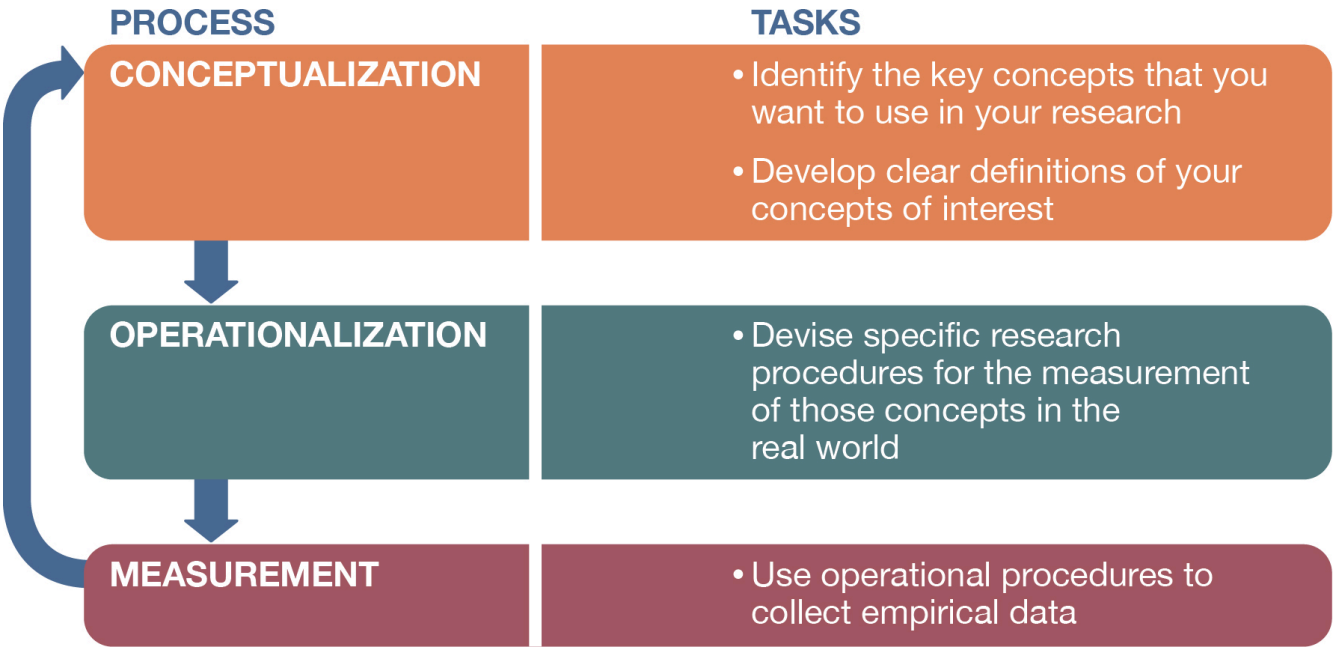


Figure 7.1. The Conceptualization-Operationalization-Measurement (COM) Process. Before you measure social phenomena in the real world, you need to do two things: (1) define all the concepts related to the phenomenon of interest (*conceptualization*); and (2) devise precise research procedures to measure those concepts (*operationalization*). After you collect the data (*measurement*), the cycle may start over again: through your analysis, you may discover that you need to develop additional concepts and operational procedures that you hadn't anticipated being important (this is especially the case for qualitative analysis).

In the process of operationalization, researchers state precisely how a concept will be measured. You can think of operationalization as the answer to the question, "What procedures will I use to measure this concept?" Specifically, you will want to **operationalize**—make choices about how to measure—any concepts that relate to your research question. Then you can go out and measure those concepts in the real world.

As we discussed in **Chapter 3: The Role of Theory in Research**, operationalization is about converting our *concepts* into measurable *variables*. (Confusingly, many social scientists use the term "variables" when they really mean "concepts"; as we noted earlier in the textbook, "variable" has become a shorthand for both conceptual *and* operational definitions.) Variables can also be called **indicators** or **measures**. Qualitative researchers rarely use the terms "variable" or "indicator," which are associated more with quantitative analyses, but at the end of the day they, too, are interested in finding ways to describe abstract concepts as they manifest themselves in the real world.

In this chapter, we'll discuss the procedures of operationalization and measurement and some of their pitfalls. As you dive into this work, one thing you should keep in mind is that your strategy for data collection will help determine your strategy for operationalization. A survey implies one way of measuring concepts, while conducting a focus group or ethnographically observing people's behavior requires quite different ways of measuring them. Your choice of a sample (what we discussed in **Chapter 6: Sampling**) will also influence your decisions about what sort of measures you can reasonably employ.

Another thing to keep in mind is that you may need to adapt and revise your measures as you learn more. The *conceptualization-operationalization-measurement* (COM) process we've described usually follows that order, with the earlier stages influencing those that come later. Sometimes, however, you'll find yourself moving back and forth between these steps. For instance, you may return to your research question and perhaps even its concepts as you think through what you can feasibly measure in the real world. After you collect data, the findings from your analysis may suggest to you that you need to refine your concepts—and the entire process may need to begin again. This is especially the case for qualitative research, in which the reworking of both concepts and measures is often the point of the study. For instance, you might want to learn how members of a group actually understand a concept, or how best to measure it in the real world, which means the measures you start with might not be those you end up with. While revising measures after data collection is more problematic in quantitative research, it is sometimes called for, too. As we've emphasized, conducting social research is not always linear or straightforward—and that goes for how we measure the social world as well.

Opening chapter image credit: NASA, *via* NASA *Earth Observatory*. *Adapted by Bizhan Khodabandeh.*

7.1. Operationalizing a Concept



What does an abstract concept like "masculinity" actually mean? In defining it, should we go with what mass media suggests to us is "masculine"—such as the buff male heroes of many action movies? Does it matter for our study that different cultures may come to very different conclusions about what it means to be "manly"? And with any particular definition of masculinity, what social characteristics, social roles, values, interests, and behavior would we focus on in determining where a particular individual falls on our measure? Whichever definitions and methods of measurement you choose (and numerous ones would be valid), you as a researcher must be prepared to defend them as legitimate. *20th Century Fox, via Wikimedia Commons*

Learning Objectives

1. Understand how the process of operationalizing a concept works.
2. Discuss how conceptualization and operationalization are related to one another.

So how do we actually go about operationalizing a concept we're interested in? We do so by identifying specific

measures that we believe can accurately represent those ideas. Let's say we're interested in studying whether people who identify as males actually exhibit characteristics of "masculinity." And let's say we conceptualize "masculinity" as "the characteristics, social roles, values, interests, and behaviors typically associated with males in a particular society." Here are some ways we might actually go about measuring masculinity (as defined in this specific way) in the real world:

1. We survey a random sample of men to gauge their levels of masculinity. We ask what kinds of movies these men like to watch. If half or more of a respondent's listed titles are action movies, we consider him to be exhibiting masculinity.
2. We listen to informal conversations in office settings. If any of the men we observe talks about contact sports, we consider him to be exhibiting masculinity.
3. We survey households and ask married men whether they are the breadwinner (defined in this case as the sole or highest-paid earner in their households). If a respondent says he is, we consider him to be exhibiting masculinity.

These are just three rough examples to give you a taste of the possibilities. We want to emphasize that the approaches to operationalization you can take as a researcher are often limitless—which is part of the fun and creativity of the research process. (In fact, some sociologists make a name for themselves precisely by figuring out ingenious ways to operationalize hard-to-measure concepts.)

You should also keep in mind that there are direct and indirect ways to measure concepts. For instance, we might explicitly ask respondents about whether they feel they are "masculine." Or, we might choose to pose questions that never mention "masculine," "feminine," or other concepts clearly related to our research question, and yet allow us to draw conclusions about how "masculine" that person would appear to others—for instance, whether the individual is a breadwinner or father or takes on some other role traditionally associated with men. A researcher could make a case that either of these approaches does a good job of measuring masculinity. Or, a researcher could say both of these approaches are flawed, in that they identify people as masculine who are really not masculine (or vice versa). Again, there is not just one way—not even just one *best* way—to operationalize a concept. The important thing is for you to be able to defend your operationalization decisions as legitimate.

Identifying indicators for some concepts can be pretty straightforward. To measure age, for instance, you can put a question on a survey that simply asks, "What was your age on your last birthday?" To evaluate someone's educational level, you might ask, "What is your highest educational degree?" or "What is the highest year in school you completed and got credit for?" Many variables are more complicated to measure, though. Imagine that you needed to record a person's income level. Would you be interested in this person's individual income, or the total income of all the people who lived in that person's household? Do you want to count income from social programs like Temporary Assistance for Needy Families (TANF) or the Earned Income Tax Credit (EITC)? Are you interested in their income per month, or per year? Before or after taxes? The specific approach you take will depend on your research question, whether you think your respondents could answer truthfully and accurately, and other practical and theoretical considerations.

For surveys and other quantitative methods that focus on apples-to-apples comparisons of one person's response to another's, it's all the more important to make good operationalization decisions from the outset that provide you with specific and clear measures. As you can imagine based on the examples we gave earlier, even an indicator measuring a relatively simple concept, such as educational level, can give you slightly different results depending on the exact phrasing of the question, and you generally do not wish to change the wording of a survey question once you start collecting data. Furthermore, for particularly complex concepts, you probably don't want to rely on just one measure. For instance, to measure masculinity and femininity, Sandra Bem (1974) selected a series of personality traits generally recognized as masculine, feminine, or neutral and asked respondents to indicate the extent to which those traits applied to them. (Check out this link to see the list of traits and try out the survey.) In **Chapter 13: Surveys**, we'll talk more about this strategy of using multiple measures to get at different dimensions of a concept.

Sometimes, a researcher will decide to ask about an aspect of a concept that is easier to measure, even if that indicator

is limited in some way. For instance, a person's precise political beliefs can be tricky to pin down, but what political party they usually support is fairly simple to find out. In the United States, a good indicator of this concept would be a survey question asking whether the respondent identifies as a Democrat, a Republican, a supporter of some other party, or an independent.[1]

The process of coming up with measures should not be too arbitrary or casual. As we mentioned earlier, one way to avoid this danger is to turn to prior theoretical and empirical work in your area of study. Theories will point you in the direction of relevant concepts and possible indicators. Empirical work will give you some specific examples of how the important concepts in an area have been measured in the past. Indeed, it often makes sense to use the same measures as researchers who have come before you. So long as you give proper credit, "borrowing" in this way in the social sciences is not only fine, but encouraged. Using a well-established measure of some phenomenon may give you more confidence in the soundness of your results. It can also give you the chance to easily compare your study's findings with those of previous studies. That said, you may notice flaws or weaknesses in measures that have been used in the past. When that happens, you can contribute significantly to the literature just by developing and deploying better measures in your own work.

Tricky Measurement Decisions

It's worth spending some time thinking about how our decisions about measurement can shape the ways we think about a particular phenomenon. You get what you measure, and if some aspect of a concept isn't captured by your measures, then it might as well not exist in your analysis of it—which can create gaping blind spots in your understanding of social life. A pretty obvious example is standardized testing in schools. Such measures (sometimes called *metrics* in nonacademic settings) understandably become an obsession for educators, given that scores on standardized tests can determine whether teachers get fired or promoted, or schools get funded or shut down. But prioritizing this particular set of metrics can come, critics say, at the expense of nurturing the types of cognitive skills (such as creative and critical thinking) that are harder to test.



Invariably whenever you hear about how the national economy is doing in the news, people will bring up gross domestic product (GDP), the combined monetary value of all the goods and services produced in the country. But using GDP as a measure of social progress can be misleading. An economy can experience strong growth and still be plagued by pollution, inequality, ill health, crime, and other negative outcomes—and, in fact, some of the costs of dealing with those problems may contribute in the short term to GDP. *Xavier Messina, via Pexels*

A particularly vivid example of how we get what we measure—perhaps for the worse—is the widespread use of the economic indicator known as *gross domestic product* (GDP). To measure the size of the national economy, government economists routinely calculate this figure, the total monetary amount of all the goods and services produced within a country's borders, usually

1. Even here, however, what might seem to be a straightforward question is not: as it turns out, most people who say they are "independent" tend to vote for one party—sometimes overwhelmingly so—even though they don't personally identify as a Democrat, Republican, or something else (Laloggia 2019).

during a year. GDP is widely cited in news reports, and political leaders brag whenever the economy is growing according to this measure (or get attacked when it is shrinking). But using GDP as a gauge of social progress has long been controversial—as the late U.S. senator Robert Kennedy famously said, "GDP measures everything, except that which makes life worthwhile." One critique of GDP is that the economy can still be growing amid increases in pollution, crime, poor health, economic inequality, and other things that we don't typically see as contributing to the wealth and well-being of society, but that nonetheless generate more market activity (oftentimes, to clean up those messes) in the short term.[2] Using GDP to chart a country's progress may therefore be overly narrow and short-sighted, giving us incentives to adopt certain policies and promote certain industries that are actually bad for society.

Let's take another example of a controversial indicator. Intelligence is an abstract concept, but we have a sense that it exists in real life—that some people are smarter than others. Many ways of measuring a person's intelligence exist, which means we have numerous ways to operationalize this concept—and great debate over which approach is best, or whether any form of rigorous measurement is even possible. One common way to measure intelligence is through tests of people's analytical and pattern-matching skills, which take the proportion of right and wrong answers to derive an IQ (intelligence quotient) score—standardized so that the average person scores 100 and two-thirds of the population score between 85 and 115. In this case, IQ is a variable that measures the concept of intelligence; it is one *operational* definition of intelligence. IQ is used throughout the world for this purpose, although some social scientists have raised concerns about how well it actually captures the multidimensional construct of intelligence—or whether such a concept of general intelligence actually exists (see the sidebar ***The Bell Curve* and Its Critics**).

A third difficult-to-measure concept is poverty. Poverty scholar John Iceland (2013:40) defines poverty as economic deprivation, but he also points out that "views about what, precisely, constitutes poverty and who the poor are still vary across both time and place." "Economic deprivation" represents a *conceptual* definition of poverty, but when we begin to think about how exactly to identify poverty in reality, we're going to have to be a lot more specific about how we distinguish a "poor" person from a "not poor" person.

Let's work through this example a bit more. If you wanted to identify "poor" people, what procedures would you use? Consider the following possibilities:

1. Interview people and ask them, "Do you consider yourself to be financially poor?" If they say "yes," then the person is poor; if they say "no," then the person is not poor.
2. Set a reasonable income threshold that one person needs to reach to achieve a basic standard of living—that income level (let's call it X) will be your poverty line, and anyone below it will be designated as poor. Interview people and ask them, "Is your yearly income before taxes below X dollars?" If they say "yes," then the person is poor.[3]
3. Interview people and ask them whether they have a working stove, a refrigerator, a washing machine and dryer, and air conditioning in the place where they live. If they answer "yes" to all these questions, then the person is not poor.

All of these methods—asking people whether they're "poor," measuring their income, and tallying their household amenities—are defensible ways that you could distinguish "poor" from "not poor" people. They each have their strengths and weaknesses. For example, a person in an affluent neighborhood may describe themselves as "poor" on a survey

2. For more background on critiques of GDP as a measure of economic progress, see Stiglitz (2020).

3. Note that when social scientists measure poverty by income, they typically look at the incomes of *households*—all the people who live together—rather than individuals. This is another reason that doing a thorough literature review is important; you should figure out what the standard approach in your area of study is.

because they make less than the ultrarich families they live near, but is that really what we mean by "poor"? Using income is the typical way that social scientists measure poverty, but some argue that a better measure is not how much people earn, but what sort of lifestyle they are able to live—which may be more accurately discerned by asking them about their access to household appliances like refrigerators and washing machines. Yet we may also miss something by focusing only on the possessions people have, which may vary by social context and doesn't account for other considerations like the amount of debt they've accumulated in the process.

At this point, you might be wondering why you even need to bother creating your own measure of poverty—doesn't the United States have an "official" measure of it? Well, yes. Let's take a quick look at the method that the federal government uses to count poor Americans, which will give you a concrete sense of just how tricky the task of social scientific measurement can be.

In 1964, President Lyndon Johnson announced that his administration was launching a War on Poverty. Drawing from preliminary ideas put forward by the Kennedy administration, Johnson and his congressional allies pushed through an array of policies to reduce the number of poor people across the country, including Head Start, Medicare, Medicaid, the Supplemental Nutrition Assistance Program (SNAP), and the Job Corps, among others. There was a problem, however: How would the government track the success or failure of these new programs? It needed a good count of the country's "poor" people over time to see whether its policies were actually effective in shrinking that population.



In 1965, Mollie Orshansky, an economist working for the U.S. Social Security Administration, came up with an influential calculation for determining who is "poor" and thus eligible for certain government benefits. The federal government adopted this poverty threshold as its official measure in 1969. Critics say it relies on a rigid, outdated standard and fails to account for the fact that society generally treats people as poor based not on whether they meet an absolute standard but rather on how they are doing relative to others. *Social Security Administration History Archives, via Wikimedia Commons*

Although various measures of poverty had been used over the years by government agencies, there was no one official measure. Then, in a series of articles, an economist in the Social Security Administration proposed using a simple and straightforward measure of poverty based on relevant data from the Department of Agriculture (Orshansky 1965). That agency had devised a number of food plans that specified how much it would cost every month to feed a family at adequate, but varying, levels of nutrition. The economist, Mollie Orshansky, took the figure for the "economy" plan—the cheapest one—and multiplied it by three to account for other necessities, such as clothing and shelter. She also calculated a separate monthly "poverty threshold" (sometimes called the "poverty line") based on the number of people in each family (Fisher 2008). In 1965, a key federal agency at the time, the Office of Economic Opportunity, started using Orshansky's procedure. Four years later, it was adopted as the federal government's official poverty measure. Even though much has changed since Orshansky devised her method (for one thing, today's households typically spend much less than a third of their budget on food), the dollar thresholds that she specified for different family sizes are still used today—adjusted only for inflation. (In 2022, the poverty line was $13,590 in annual income, before taxes, for one person.)

Is using Orshansky's poverty thresholds the "best" way to measure whether or not a person is poor? According to critics, the official measure is based on outdated notions of how much a household's necessities cost. It fails to adjust for geographic differences and ignores the noncash government benefits people receive or the wealth they hold. Alternative measures of poverty that address these concerns have gained the interest of policymakers, but Orshansky's measure continues to be the key number touted in news headlines and used to make important bureaucratic decisions about who receives assistance.

Even if the problems we just noted were addressed, some social scientists argue that the official measure would remain fundamentally flawed because it is an *absolute* measure: the standard that it uses—the cost of an economy food plan decades ago—stays the same, other than being adjusted for inflation, and therefore it doesn't account for dramatic changes in lifestyles and living standards over the years. In other countries, governments track their progress by instead measuring *relative* poverty—comparing how much worse people are doing relative to the typical, or median, household. (For example, many countries in the European Union identify poor people as those households earning under 60 percent of the country's median income.) This relative measure may better reflect the psychological reality of being poor—that people actually judge whether they are rich or poor based not on how much they make per se, but whether they make more or less than a relevant **reference group**—as the saying goes, whether they are "keeping up with the Joneses." Other social scientists argue, however, that an absolute measure is superior because (regardless of the envy that some people

feel for the better-off) society as a whole shouldn't really care about who has more money, but rather whether everyone has a decent standard of living—which was the point of Orshansky's original formula.

Who's right? Clearly, measurement in sociology, especially measurement of abstract concepts like "poverty," is a complex process involving many decisions by researchers, some easier to defend than others, and some inherently political in nature. As you figure out how to measure the concepts you're interested in, keep in mind that social measures are almost never perfect and should always be open to rethinking and revision. Operationalization is probably one of the trickiest components of basic research methods. Don't get frustrated if it takes a few drafts and a lot of feedback to get to a workable specification of your measurement procedures. And be sure to pilot test, pilot test, pilot test!

Conceptualization and Operationalization in Qualitative Research

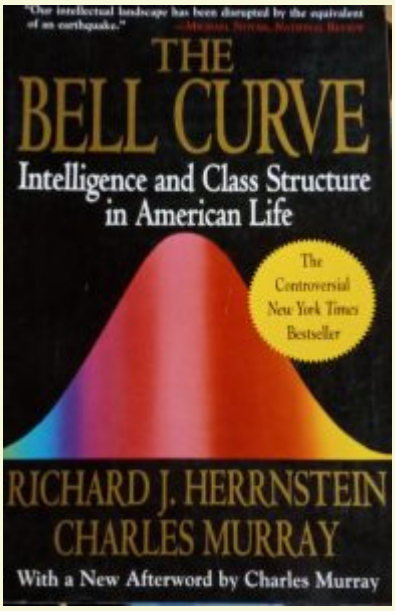
If you were a quantitative researcher studying masculinity, you might use a battery of survey questions like the Bem Sex Role Inventory to operationalize this concept in your study. If you were doing in-depth interviews, however, the strategy you adopt would probably be very different. Qualitative research takes a more open approach toward defining the concepts relevant to a particular study, and a qualitative interview is much less structured than a survey interview. You might start off with a tentative definition of contemporary masculinity based on the past literature. You could then use that working definition to come up with an initial set of questions for interviewees. After talking to people, you might add questions—on the fly during an interview or afterward to prepare for the next interview—to refine or expand your understanding of the relevant concepts.

The ways that qualitative sociologists define their concepts and operationalize them often evolve over the course of their research. When conducting in-depth interviews, for example, they may ask a set of recurring questions to measure particular concepts of interest, but they will probe more deeply into areas of interest and even change their list of questions as they gather more information. Sociologists employing ethnographic observation may start with certain definitions in mind as they categorize the behaviors they see in the field, but they will try to refine those conceptualizations with every new observation. Ultimately, the concepts in a qualitative study will be defined by the researcher's interpretation of what interviewees say or what the researcher observed, and their ways of measuring those concepts may change in line with that shifting understanding.

As a qualitative researcher, you should be prepared for the degree of improvisation and adaptation required in this work of continually reconceptualizing and reoperationalizing whatever you are studying. As we will elaborate on in the later chapters devoted to qualitative methods, you typically want to seek out new ideas and possibilities that emerge from your fieldwork rather than sticking stubbornly to any set theoretical playbook you had when you started your study.

The Bell Curve and Its Critics, by Andrew Park

In their controversial book *The Bell Curve: Intelligence and Class Structure in American Life* (1994), psychologist Richard J. Herrnstein and political scientist Charles A. Murray described an America that was increasingly divided between a wealthy, well-educated elite and an impoverished underclass. That view of inequality is actually in line with the view of many sociologists today, who in recent decades have written extensively about the country's growing economic and cultural divides. What made *The Bell Curve* so inflammatory was what it argued was driving this rising inequality: genetically based differences in intelligence. Methodological problems in Herrnstein and Murray's research drew widespread criticism at the time—including from prominent sociologists—yet three decades later, the book continues to find audiences supportive of its arguments.



The Bell Curve (1994) argued that differences in IQ—including racial differences—were decisive in explaining the outcomes of people's lives, a claim that drew harsh criticism, including from a team of sociologists who wrote a book-length rejoinder, *Inequality by Design* (1996). *Simon & Schuster and Princeton University Press*

The Bell Curve argued that people's life trajectories in terms of earnings and well-being were better explained by their measured levels of intelligence, rather than their family background, and that intelligence was largely an inherited trait fixed at birth. Herrnstein and Murray analyzed the responses of thousands of respondents in the National Longitudinal Survey of Youth (NLSY), a panel study that tracks the same individuals over many years. They set out to discover whether any correlation existed between the respondents' scores on an IQ test and various negative outcomes later in their lives, such as becoming poor, unemployed, or a high school dropout. Furthermore, they wanted to evaluate whether IQ scores were a better predictor of respondents' outcomes than the socioeconomic status of their parents. They concluded that IQ scores were more decisive—for instance, 30 percent of the respondents who scored in the bottom 5 percent of the IQ test ended up poor, whereas only 24 percent of those whose parents were poor ended up poor as well. Herrnstein and Murray also noted racial differences in IQ scores, which they suggested were also tied to genetic differences.

In Herrnstein and Murray's view, the stronger influence of IQ scores on life outcomes meant that any existing racial and class inequalities were more the result of differences in actual ability than discrimination or a lack of opportunity. Indeed, Herrnstein and Murray took their findings to mean that social policies to improve access to educational and other opportunities for less advantaged groups were doomed to failure—what ultimately mattered in determining people's later life outcomes was the genes passed down to them, not the environment they grew up in.

A team of sociologists—Claude S. Fischer, Michael Hout, Martín Sánchez Jankowski, Samuel R. Lucas, Ann Swidler, and Kim Voss—published a book, *Inequality by Design* (1996), which critiqued Herrnstein and Murray's work on a number of levels. First, although Herrnstein and Murray defined intelligence as "a general capacity for inferring and applying relationships drawn from experience," the sociologists questioned any broad conceptualization of "general intelligence," given how many different dimensions it can conceivably

cover—from mental quickness, to problem-solving ability, to creativity—with some of these forms of intelligence not necessarily related to the others.

Even if we assume there is such a thing as general intelligence, the way the *Bell Curve* operationalized this concept was flawed as well, the sociologists wrote. The IQ test referenced in *The Bell Curve* is the Armed Forces Qualifying Test (AFQT), a test that is used to determine enlistment eligibility and is composed of sections evaluating arithmetic reasoning, mathematics knowledge, paragraph comprehension, and word knowledge. But Fischer and his collaborators argued that the AFQT really measures "schooling," rather than "intelligence"—it tests basic knowledge and foundational skills taught in primary school or in children's home environments. If students from better-off households have access to more effective schooling or more nurturing home environments, then the test is at least partly measuring their social advantages, not just their underlying smarts. For a similar reason, the low IQ scores of African Americans and Hispanics relative to whites and Asian Americans really demonstrate their low social status rather than any genetic differences, the sociologists concluded. Any subjugated group—whether Korean minorities in Japan or Polish Jews in early twentieth-century America—comes to be stereotyped as innately unintelligent, they pointed out; once a group's relative social status improves—and thus their access to opportunities widens—those "intelligence" inequalities disappear.

Finally, Fischer and his fellow sociologists took issue with Herrnstein and Murray's actual empirical analysis. The way that Herrnstein and Murray operationalized environmental factors was much too blunt and narrow, they argued. Beyond the socioeconomic status of parents, other variables—such as parental education, geographic location, and number of siblings—needed to be controlled for; otherwise, any measured correlation between people's IQ scores and their later life trajectories might actually be masking the influence of those confounders. When such environmental factors were accounted for, the measured effect of IQ scores on the risk of one day becoming poor or experiencing other negative outcomes dwindled away, the sociologists found.

Fischer and his colleagues asked to debate Murray on the merits of the *Bell Curve*, but he has never taken them up on the offer (his coauthor Herrnstein died in 1994). Almost three decades later, though, the *Bell Curve* continues to draw a following, with Murray's 2017 unapologetic defense of the *Bell Curve* on a popular podcast unleashing another wave of debate over it. This time, psychologists Eric Turkheimer, Kathryn Paige Harden, and Richard E. Nisbett weighed in. Unlike Fischer and his collaborators, the psychologists—who work regularly with intelligence tests in their own research—were less skeptical about such kinds of testing, which they point out "can be meaningful and useful even if an essential inner [general intelligence] doesn't exist at all." However, recent research has only underscored, they said, the point that IQ is not a fixed quantity but rather responds markedly to environmental factors. Studies find that children adopted from poor homes into affluent ones experience 12- to 18-point IQ gains. Test score gaps between African Americans and whites have shrunk over the decades. In a phenomenon known as the "Flynn effect" (named for the IQ researcher James Flynn), IQ has increased dramatically in many countries in recent decades—by 18 points in the United States from 1948 to 2002.

These findings all suggest that changes in people's social environments can significantly affect their IQ scores. More broadly, they make us question whether the ways we social scientists routinely operationalize intelligence truly capture the underlying concept—assuming that there really is such a thing as "general intelligence" to begin with.

Key Takeaways

1. The typical steps in the conceptualization-operationalization-measurement process in quantitative research involve developing a research question, defining the key concepts related to that question, and deciding on operational definitions of those concepts—that is, how exactly to measure them in the real world.
2. Using overly narrow measures can lead to blind spots in our understanding of social life. We can focus just on those metrics and ignore other things that may be more important or actually a better gauge of the underlying concept we're interested in.
3. The measurement process in quantitative and qualitative research differs in that qualitative research often involves the continual revising of procedures and operational definitions as more data is collected.

7.2. Deciding on the Correct Level of Measurement

Learning Objectives

1. Learn how the four levels of measurement can be distinguished from one another.
2. Describe why mutual exclusivity and collective exhaustiveness are important in the operationalization of variables.

When social scientists measure concepts, they often use the language of variables and attributes. In this context, you can think of a *variable* as a grouping of two or more characteristics, and **attributes** as the characteristics that represent values or categories of a variable. For example, the variable "dog breed" would contain attributes like terrier, retriever, and shepherd. How many attributes a variable has, and how those attributes can be distinguished from one another, are especially important in quantitative research. A variable's attributes determine its **level of measurement**, which in turn tells us what kinds of mathematical operations can be applied to the variable when conducting a quantitative analysis.

For practical purposes, the three possible levels of measurement in sociology are nominal, ordinal, and scale. (See the sidebar **Deeper Dive: Interval and Ratio Levels of Measurement** to learn about even more granular levels of measurement.) A variable at the **nominal level** has attributes that are different from one another, but unlike for the other levels of measurement, those attributes do not follow any mathematical order.

Dog breed is a good example of a nominal-level variable. In 2022, the American Kennel Club (AKC) recognized 199 dog breeds. Each breed's ideal physical traits, movement, and temperament are carefully described in a "breed standard," which specifies its unique characteristics. **Figure 7.2** provides an example of two dog breeds whose characteristics are different from one another.



Figure 7.2. Nominal Level of Measurement Example: Dog Breeds. We can consider dog breed to be a nominal-level variable. Two of its (numerous) attributes are German shepherd and terrier. These two breeds have different characteristics and qualities. One is not "more than" or "less than" the other—they are just different. That means we should classify this variable as nominal level. *Summer Stock and Sam Lion, via Pexels*

To identify a nominal-level variable, ask yourself one question: "Can I rank the variable's attributes in order from high to low (or low to high)?" The ranking has to make logical sense. For example, a higher-ranked attribute must be higher in some characteristic than the ones below it, such as a higher frequency of occurrence, higher amount of time spent, higher level of ability, or higher degree of satisfaction. If your answer to this key question is "No"—the variable's attributes cannot be ranked—then the variable is nominal level. If your answer is "Yes," then the variable should be measured at a higher level.

The answer to this question for our "dog breed" variable is clearly "No": the diverse categories of dogs captured by the variable do not have any inherent ordering to them. That said, two conditions must hold for this nominal-level variable—or, for that matter, any variable we want to analyze quantitatively. First, its attributes must be **mutually exclusive**: the categories must all be different from one another, with no overlaps. Second, its attributes must be **collectively exhaustive**: the categories should cover all the conceivable instances that might occur—that is, they should "exhaust" all possibilities. (Sometimes these two conditions are abbreviated as MECE, or "mee-cee": mutually exclusive + collectively exhaustive.)

How do you ensure that your nominal-level variable is exhaustive? An easy way to do so on surveys is to allow respondents to answer "none of the above" or "other" in answer to the question. That way, the survey question's **response categories**—the answers available for each question—will truly cover all possibilities (we'll have more to say about survey question design in **Chapter 13: Surveys**). To make our dog breed variable exhaustive, we could create two "other" categories: an "other breed" attribute to cover dogs who are purebred, but not of a type recognized by the AKC, and a "mutt" attribute to cover dogs of mixed or unknown breeds.

It can be trickier at times to ensure that your variable's attributes are mutually exclusive. In plenty of situations, "overlaps" across categories are important. For instance, if you are measuring race, you'd want to account for the large number of people with mixed-race identities. Creating an "other" category is the simplest way of addressing this issue, but you may also want to create separate categories for common multiple-race identities. (You can also create

multiple variables to preserve information about cases that straddle multiple categories, as we discuss in **Chapter 14: Quantitative Data Analysis**.)

Following are some examples of nominal-level variables that are commonly used in sociological research. Each of these variables has a range of attributes that are mutually exclusive and collectively exhaustive.

Table 7.1. Sociological Examples of a Nominal Level of Measurement

Variable	**Attributes**
Gender	1. Female 2. Male 3. Nonbinary 4. Other (*please specify*)
Religious affiliation	1. Catholic 2. Jewish 3. Protestant 4. Other religion (*please specify*) 5. Not affiliated
Political affiliation	1. Democrat 2. Republican 3. Other party (*please specify*) 4. Independent
Sexual orientation	1. Heterosexual or straight 2. Gay 3. Lesbian 4. Bisexual 5. Other (*please specify*)
Employment status	1. Employed full time 2. Employed part time 3. Retired 4. Not employed for pay
Country of residence	1. Afghanistan 2. Albania 3. Algeria 4. Andorra . . . to Zimbabwe

In **Table 7.1**, we've numbered the responses—something you will want to do whenever you input survey data into a

quantitative dataset.[1] Note that although the attributes in this table are numbered, the order of the attributes for nominal-level variables is always arbitrary. Protestants, for example, could have been numbered as 1, 2, 4, 5, 103, or any other unique number. (Sometimes surveys take advantage of their flexibility to number response categories by grouping together similar categories in a number range—for instance, making Catholics 10, but putting Protestant denominations in the thirties, such as 31 for Lutherans, 32 for Methodists, and so on.)

After nominal-level variables, the next higher level of measurement is the **ordinal level**. Like nominal-level variables, ordinal-level variables must have attributes that are mutually exclusive and collectively exhaustive. Unlike nominal-level variables, they have attributes that can be ranked from high to low (or from low to high) using some kind of meaningful comparison. For example, we could ask survey respondents to rate how healthy they feel from "poor" to "excellent." We would then have an ordinal-level measure of perceived health.

Ordinal-level variables are very common in quantitative sociological research. In the surveys you've taken, you've probably come across **Likert scales**, which are used to measure the intensity of people's opinions. (By the way, it's pronounced *LICK-ert*—something that many professional sociologists get wrong!) To measure agreement with a particular statement, Likert scales will often use four or five ranked response categories that range from "Strongly agree" to "Strongly disagree." Consider this question asked in the General Social Survey, a survey of the attitudes of U.S. adults taken every two years:

> How much do you agree or disagree with the following statements? American culture is generally undermined by immigrants.
>
> 1. Strongly agree
> 2. Agree
> 3. Neither agree nor disagree
> 4. Disagree
> 5. Strongly disagree

In this example, "strongly agree" (numbered as 1) represents the most positive attitude toward the statement, while "strongly disagree" (5) represents the most negative attitude toward the statement. "Neither agree nor disagree" (3) is an intermediate category, representing a neutral or undecided attitude.

The only thing that matters when we number ordinal-level variables is the ranking, not the specific number assigned. For instance, we could easily switch 1 to mean "strongly disagree" and 5 to mean "strongly agree." Under this new numbering, increasing values would mean greater agreement with the statement; in the original, they mean greater *dis*agreement. Also, the response categories do not have to be numbered from 1 to 5—we could number them 0 to 4, or −2 to +2, for instance. For ordinal-level variables, the numbering could be reversed or be different numbers entirely, as long as the numbers are still ordered in relation to each other.[2]

1. Among other things, numbering your response categories makes it easier to enter your data into a statistical analysis software package. For instance, instead of typing out "female" for every person who identifies as such in your dataset, you could just enter "1" for the gender variable. This avoids the danger of mistyping "female" into your dataset, which might prompt your statistical analysis program to classify that particular person under a different response category—such as when the program mistakes "Female" as a different category from "female."
2. On some occasions, you may want a specific numbering for an ordinal-level variable. For instance, class rank—an

Ordinal-level variables differ from higher-level variables in one key way: although we can rank the attributes of ordinal-level variables, we cannot determine the exact distance between those ranks. **Figure 7.3** visualizes what we mean by "distance" using the example we previously mentioned: level of perceived health.

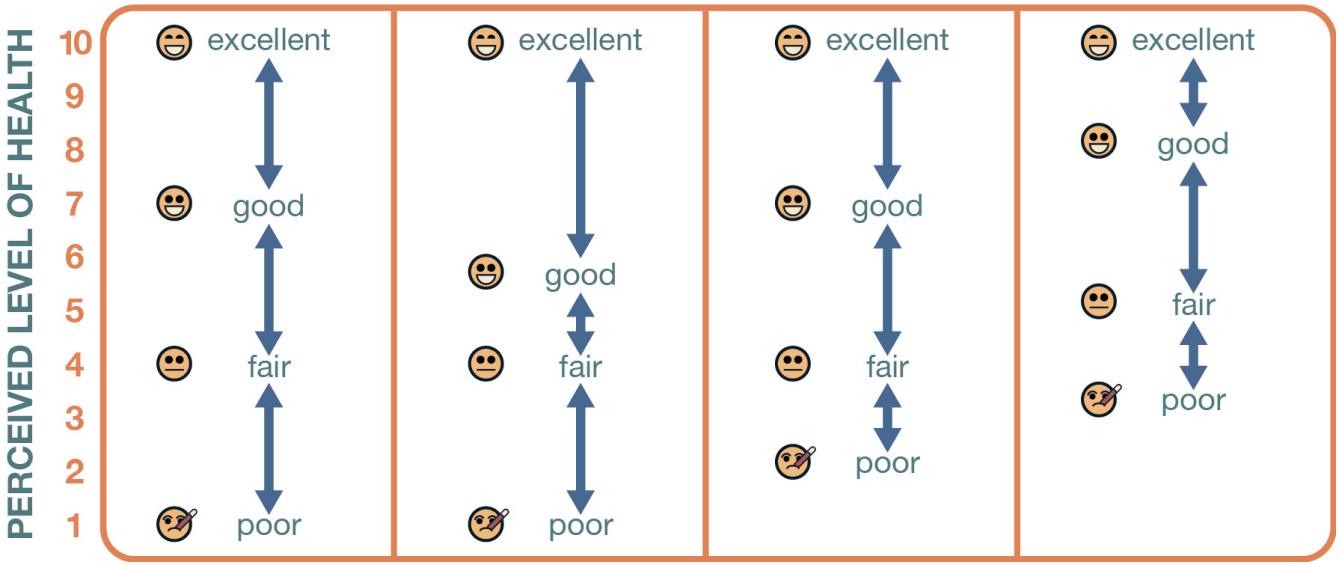


Figure 7.3. Ordinal Level of Measurement Example: Perceived Health. This figure shows us four ways that an ordinal-level variable—with response categories of excellent, good, fair, and poor—could map onto a scale of 1 to 10. Which set of distances between the response categories is correct? With an ordinal-level variable, we simply don't know. Although the relative positions of the four categories are clear, their absolute positions—just how healthy "good" is on a 1–10 scale, for example—are unknown. Furthermore, the distance between the categories could be interpreted differently by different respondents based on what they think the terms (e.g., "fair") actually mean.

For our variable of each respondent's perceived level of health, the attributes are *excellent*, *good*, *fair*, and *poor*. Let's assume we actually know how healthy each person is on a scale of 1 to 10 scale, with 10 being the healthiest (for various reasons, this is simplistic, but let's go with it!). How do the response categories for our ordinal-level variable of perceived health map onto that scale? We know that "excellent" represents the highest level of health and "poor" the lowest, but we do not know where other responses would put us on our "real" 1–10 scale—is a response of "good" a 6, 7, or 8? It's possible that each respondent interprets words like "good" or "fair" in different ways. One person may think of "good" as the equivalent of 8 on the 1–10 scale, whereas another person thinks of it as a 6. Likewise, we know that "good" means worse health than "excellent," and "fair" means better health than "poor," but we don't know that whether the distance between one particular category and another is 1, 3, 5, or 7 points—or whether those categories would be spaced out equally across such a scale. All we know is the rank ordering.

ordering usually based on GPA that indicates where exactly a student falls relative to their peers—is an ordinal-level variable, but it already has a built-in numbering scheme, with "1" being the person who is at the top of the class. IQ is technically an ordinal-level variable (although some social scientists treat it as higher level), given that it measures a person's intelligence relative to the larger population; as discussed in the sidebar ***The Bell Curve* and Its Critics**, it also has its own numbering scheme, which sets an IQ of 100 as the population median, and quantifies differences between individuals so that two-thirds of the population scores between 85 and 115.

Table 7.2 provides more examples of common ordinal-level variables in sociology. As you review these examples, think about the distances between the response categories and whether they do, or do not, change as you move across the continuum. Also think about what the numbering means—for example, as the numbers rise, what exactly is increasing?

Table 7.2. Sociological Examples of an Ordinal Level of Measurement

Variable	Attributes
Confidence in U.S. Congress	1. A great deal 2. Only some 3. Hardly any
Highest educational degree	1. Less than high school diploma 2. High school diploma 3. Some college completed 4. Bachelor's degree 5. Graduate degree
Difficulty of giving up cell phone	1. Impossible 2. Very hard 3. Somewhat hard 4. Not too hard 5. Not hard at all
Satisfaction with family life	1. Very satisfied 2. Somewhat satisfied 3. Somewhat dissatisfied 4. Very dissatisfied
Frequency of religious service attendance over the past year	1. More than once a week 2. Once a week 3. Once or twice a month 4. A few times a year 5. Less than a few times a year 6. Never
Family income before taxes last year	1. Less than $20,000 2. $20,001–$50,000 3. $50,001–$75,000 4. $75,001–$100,000 5. $100,001 or more

You might have noticed that two of the variables in this table—frequency of religious service attendance and family income—could, in theory, be measured more precisely. It's often difficult, however, to obtain precise and accurate values for these variables. Do people remember how much money they made in dollars before taxes last year, or exactly how many times they went to religious services last year? Maybe some do, but many others will be stumped. Particularly

for income, people may feel uncomfortable giving you a specific number. Such sensitive or difficult-to-recall variables are often collapsed into ordinal groups, so that we can tell whether one person has a higher level than another, but not exactly how much higher.

To summarize what we've discussed so far, nominal-level variables must have attributes that are *mutually exclusive* (no overlaps) and *collectively exhaustive* (all possible responses accounted for). Ordinal-level variables add another condition: *rank ordering* (the attributes correspond to higher or lower ranks). The next-highest level, the **scale level**, meets these three conditions and adds one more: *equal intervals*. You can see the relationships among these three levels of measurement visualized in the hierarchy shown in **Figure 7.4**, which goes from the variables with the least to those with the most conditions.

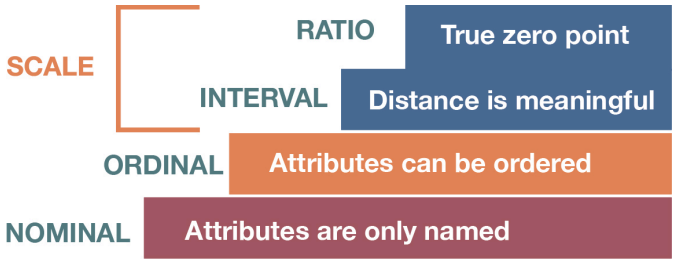


Figure 7.4. The Hierarchy of Levels of Measurement. This figure summarizes the different levels of measurement and their characteristics. Any measure fulfilling the conditions of mutual exclusivity, exhaustiveness, rank order, and equal intervals is said to be scale level. Take away the equal-intervals condition, and the highest level is ordinal. Take away the rank-order condition, and you are left with nominal-level data. (As we describe in the sidebar Deeper Dive: Interval and Ratio Levels of Measurement, scale-level data can be further distinguished based on whether the variable has a "true zero point"—if that condition holds, the variable is ratio level, and if not, it is interval level.)

When we say that a scale-level variable has "equal intervals," we mean that the distance between each pair of adjacent ranks or points on the scale is equivalent. Social scientists often use the term *continuous variable* interchangeably with "scale-level variable," and you might want to think of the values of a scale-level variable as lying on a continuum much like a number line. Every point along that continuum has a possible value, and the distances between the adjacent ranks—on a number line, the numbers next to each other—are equal.

Perhaps the scale-level measure that you come across most often in your daily life is temperature. Both the Fahrenheit and Celsius scales meet this additional condition of equal distances on their scales measured in degrees: the difference between, say, 20 and 21 degrees is the same as the distance between 42 and 43 degrees. With an ordinal-level measure of temperature—say, "hot," "lukewarm," and "cold"—we wouldn't know if the distance between "hot" and "lukewarm" was equivalent to that between "lukewarm" and "cold." With scale-level variables, we know the distances, and they are equal across each rank (in this case, each degree of temperature).

Table 7.3 presents sociological examples of scale-level variables—which would include all the variables that we have not already categorized as nominal or ordinal level.

Table 7.3. Sociological Examples of a Scale Level of Measurement

Variable	Attributes
Years of formal education	0, 1, 2, 3, 4, and so on, up to the highest number of years mentioned
Number of siblings	0, 1, 2, 3, 4, and so on, up to the highest number of siblings mentioned
How many days the respondent attended religious services in the past week	0, 1, 2, 3, 4, 5, 6, 7
Parent's age in years at birth of first child	Lowest age mentioned (e.g., 15) up to highest age mentioned (e.g., 78)
Number of days in past month respondent has used one or more social media sites	0, 1, 2, 3, 4, 5, and so on, up to 31

Whether the distance between a variable's attributes is measurable and equal is critical because it affects the specific procedures we use to analyze the data. As we will discuss further in **Chapter 14: Quantitative Data Analysis**, we can do certain calculations—such as determining means and medians—for scale-level variables that we can't do for nominal-level and ordinal-level variables. (Variables at those lower levels of measurement are referred to as **categorical variables** to distinguish them from continuous variables.)

If you have the choice, it's generally best to collect data at as high a level of measurement as possible. That's because higher-level data can be grouped and converted into a lower level of measurement. For example, if we collect scale-level variables like an adult respondent's age at a lower level—say, in ordinal groups of 18 to 34, 35 to 49, 50 to 64, and 65 and older—we can never recover the precise ages during data analysis. If the only thing you know is that a person is in the age range of 18 to 34, it is impossible to say exactly how old they are. It could be 18, or 34, or anything in between. You would also run into problems taking the halfway mark—say, 26—to stand in for all the people who chose that response, because it may be that most of the people who answered that way fell in the lower (or upper) range of that category.

Conversely, if you know your respondents' specific ages, you can easily convert your data into ordinal groups at any point in your analysis. That said, researchers often choose not to measure some concepts at the scale level even when they can do so. As we alluded to earlier, getting respondents to recall precise figures can be difficult, and they are especially reluctant to be pinned down on specifics in regards to sensitive topics like money, sex, or drug use.

Deeper Dive: Interval and Ratio Levels of Measurement

Scientists further divide scale-level variables into interval-level measures and ratio-level measures. The critical distinction between these two levels of measurement is that ratio-level variables have a *true zero point*. Having

a “true zero point” means that zero on whatever scale is being used refers to the complete absence of the phenomenon being measured. Let’s take the example we used previously, temperature, which measures thermal energy.

For a measure of temperature to have a true zero point, a zero on the scale should refer to an environment that has utterly no thermal energy. That does not hold for the Fahrenheit or the Celsius scales. Indeed, both scales can go below zero into negative numbers (although you might prefer a wintry day with temperatures below zero in Fahrenheit to one spent below zero in Celsius). Because they do not have a true zero point, the Fahrenheit and Celsius scales are measured on the **interval level**.

This distinction has mathematical importance. The fact that Fahrenheit and Celsius do not have a true zero point means that any *ratios* of their numbers are not meaningful. For example, on both scales, a room heated at 40 degrees is not twice as hot as one heated at 20 degrees (2:1 ratio); there is not twice as much thermal energy in the first room as there is in the second. Likewise, a substance cooled down to −30 degrees is not three times as cold as one cooled to -10 degrees (3:1 ratio). To be able to say that such ratios are meaningful, we need the measure to have a zero point that actually refers to *nothing*—that is, the complete *absence* of whatever we’re measuring. For temperature, it’s the point at which there is zero thermal energy in a system—the coldest temperature possible.

As you might remember from other classes, this is actually true on the Kelvin temperature scale. (As you can see in **Figure 7.5**, Kelvin is simply Celsius adjusted upward by 273 degrees, with 273 Kelvin equivalent to 0 degrees Celsius.) Zero Kelvin is the complete absence of thermal energy. This means that the Kelvin temperature scale has a true zero point, and it is measured on the **ratio level**—the highest level of measurement.

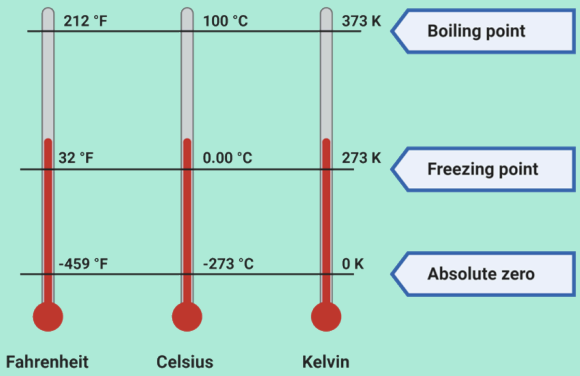


Figure 7.5. Fahrenheit, Celsius, and Kelvin Temperature Scales. Temperature scales provide a good example of the distinction between the interval and ratio levels of measurement. Fahrenheit and Celsius are interval-level because the distances between points on those scales are meaningful and represent equivalent changes—for instance, a one-degree increase from 39 to 40 degrees is the same in magnitude as a one-degree increase from 10 to 11 degrees. However, these two temperature scales are not ratio-level because both lack a true zero point: neither 0 degrees Fahrenheit nor 0 degrees Celsius represents the complete absence of what the scales are measuring (thermal energy). Zero degrees Kelvin, however, has this meaning and is therefore ratio-level. Note that all three of these temperature scales are considered to be at the scale level of measurement, which includes interval-level and ratio-level measures. *Emeka Udenze, via Wikimedia Commons*

For such variables, ratios are meaningful. Twenty degrees on the Kelvin scale is twice as hot (has twice as much thermal energy) as 10 degrees. A person who is 12 years old is six times as old as a 2-year-old. $100,000 in cash is 10 times as much as $10,000. A 20-year-old who has lived in the United States all her life has lived there four times as long as someone who moved to this country five years ago. For all these variables, the zero on the scale has a meaning, even if it is a social or cultural meaning. For example, the zero point for age is the point right before birth, the point of "zero" age (it's worth noting that some cultures have historically tacked on nine months to that age to account for time in the womb). Zero is also meaningful for a variable measuring the number of years you have spent at your current address—although whether that point of measurement begins right when you walk over the threshold of your new home or when you sign an official document is often arbitrary.

When you are doing quantitative data analysis in sociology, interval and ratio levels don't actually have any practical differences. You can use the same analysis techniques with each. In the research literature, you'll often see them both referred to as scale-level variables. When you see this term, keep in mind that this is not an additional level of measurement. Instead, it refers to both interval and ratio levels.

Key Takeaways

1. Variables measured at the nominal, ordinal, and scale level of measurement have different characteristics that affect how they can be used and interpreted in social research.
2. Mutual exclusivity and collective exhaustiveness characterize all these levels of measurement, but ordinal-level variables have rank ordering, and scale-level variables have meaningful distances between their values in addition to rank ordering.
3. It's generally a good idea to measure phenomena at the scale level, because you are stuck with the degree of precision you used when you collected your data. However, for sensitive questions or hard-to-remember figures, researchers may choose to go with a lower level of measurement.

7.3. Indexes and Scales: Measures with Multiple Items

Learning Objectives

1. Learn the two ways in which indexes and scales differ from one another.
2. Be able to provide a concrete example of an index and a scale.

Many variables can be captured by a single question on a survey. Age and income are good examples: "What is your age?" is as straightforward as you can get, and so long as we operationalize income precisely (i.e., being clear about whether we're asking about individual or household income, before taxes or after taxes), one well-worded item should suffice.

Unfortunately, most variables in the social world are not so simple. For example, it is hard to think of one item that would adequately measure a complex concept like sexism. To gauge whether a respondent has sexist views, for example, researchers might ask whether the person agrees with a statement like "It is more important to encourage boys than to encourage girls to participate in athletics," with agreement indicating that the person holds sexist views (as defined in this way). But we might be skeptical that asking this one question truly reflects the wide variety of types of sexism.

For this reason, researchers routinely use multiple indicators to develop a single measure of a complex concept. Take the concept of "well-being," for example. The Gallup organization, a well-respected polling outfit, conducts an ongoing survey of the American adult population about their well-being. The survey asks participants numerous questions covering five broad areas: career, social, financial, community, and physical well-being. Gallup uses their answers to questions about these five areas of life as indicators of the different dimensions of the complex concept they want to study.

As another example, suppose we wanted to measure how accepting the general public is of members of a particular racial or ethnic group. Because people might accept the presence of members of that group in some situations, but not others, ideally we would need to use more than one item to develop an overall measure. For example, we could ask several questions representing different levels of acceptance, such as the following:

1. Would you be willing to have a member of this group visit your country?
2. Would you be willing to have a member of this group as your colleague at work?
3. Would you be willing to marry a member of this group?

Note that this question format is a condensed version of a Bogardus Social Distance Scale, named after the sociologist who invented this technique for measuring acceptance of social groups (Bogardus 1933). The three questions can be combined into a single variable that measures how much each respondent tolerates members of the specified group.

When multiple indicators are combined into a single composite measure, the resulting variable is called an **index** or a **scale**. Researchers often use the terms "index" and "scale" interchangeably, and they are similar in that they both use two or more indicators to generate one variable that measures a single concept at an ordinal or higher level of measurement. Technically speaking, however, scales and indices (also written correctly in the plural as "indexes") differ from one another in two ways:

1. The items that comprise an index are conceptually different from one another, whereas the component items in a scale are conceptually similar.
2. The component items in a scale have some kind of intensity structure: as you move across the items, you see a greater intensity of experience, behavior, belief, and so on, being measured (such as how the Social Distance Scale moves toward more intimate relations with the members of a particular group—from tolerance of having them as tourists to tolerance of marrying them).

Let's see how indices and scales work by considering several examples. Researchers at the Social Progress Imperative, a think tank based in Washington, DC, sought to compare the progress of countries around the world in three areas: *basic human needs*, *foundations of well-being*, and *opportunity*. They developed an index with a wide range of indicators covering each of these broad concepts and their various dimensions. **Figure 7.6** diagrams the overall scheme.



Figure 7.6. Social Progress Index. The Social Progress Index combines measures that speak to different dimensions of well-being in the lives of members of a society. This list shows the 12 basic areas of social life that researchers investigated and the subdimensions of those dimensions that specific measures tracked. For example, "access to basic knowledge" is based on a calculation that includes the percentage of women in the society with no schooling at all, and "shelter" includes the percentage of deaths attributable to household air pollution. In turn, all 12 basic areas were mathematically combined into an overall index score for each country. *Social Progress Imperative*

Rather than fielding their own surveys, the think tank's researchers based their index's measures on existing data collected by national and international organizations. For example, for the Nutrition and Basic Medical Care dimension, they used data on maternal mortality rates for each country from the World Health Organization. Because survey data measuring "acceptance of gays and lesbians" was not available for many Middle Eastern countries, they substituted a measure based on legal treatment of LGBTQ+ persons in those countries. We can debate the merits of these methodological choices, but it's worth recognizing how much practical considerations can matter in the construction of any index or scale.

As is the norm for indices, the indicators used in the Social Progress Index are quite different from one another

conceptually, even though they are designed to represent an overall concept of "social progress." They are also measured in dramatically different ways. For example, "access to quality health care" in the Health and Wellness section is measured by country experts' views of the percentage of citizens who are unable to "exercise their political rights as adult citizens" because they receive health care of a poor quality. In contrast, "expected years of tertiary schooling" ("tertiary" refers to learning in colleges and universities) in the Access to Advanced Education section refers to the "number of years a person of tertiary school entrance age can expect to spend within tertiary education," based on government data.

The think tank's researchers used mathematical techniques to combine these very diverse indicators. They then generated numerical scores for each country's overall "social progress" as well as its progress within each dimension and subdimension. (You can find more details in the Social Progress Imperative's methodology report and on its data page.)

So, how exactly is a scale different from an index? Let's compare the Social Progress Index to the Bogardus Social Distance Scale we mentioned earlier. In the Bogardus scale, the indicators are conceptually similar to one another. All focus on situations involving the level of acceptance of a member of a particular social group. The first situation ("visit your country") represents the lowest level of acceptance of the three items, and the third ("marry") represents the highest level of acceptance. When compared with one another, then, the indicators form a logical intensity structure. If we scored people's responses as "No" = 0 and "Yes" = 1, we would likely see the following patterns:

Table 7.4. Acceptance of Social Groups Scale

Acceptance of Social Groups (0 = No, 1 = Yes)				
Visit your country	0	1	1	1
Colleague at work	0	0	1	1
Marry	0	0	0	1
Total score	0	1	2	3

Note: This scale is based on the Bogardus Social Distance Scale.

Table 7.4 shows the patterns of responses that the "acceptance of social groups" scale will most likely generate. (It is based on the Bogardus Social Distance Scale, although the original scale created by sociologist Emory S. Bogardus had more questions and a different scoring system.) A respondent who has a total score of 0 obviously rejects this group in all three situations. A person with a score of 3 accepts this group in all three situations. But consider a respondent who has a total score of 1. That person is likely to be willing to have the group visit their country, but not be willing to accept the more intimate situations. It is possible that someone with a score of 1 could have answered "Yes" to the most intense level of acceptance—being willing to marry someone of this group—and then "No" to one or both of the other scenarios, but that would not make much sense because agreement with the last item implies agreement with all the preceding items. For the same reason, a respondent with a score of 2 is likely to accept the two less intimate scenarios but not the third one; any other set of responses that add up to 2 would not fit the logic of the questions. Because the component indicators have this logical intensity structure, they form a scale that measures acceptance of varying levels of intensity of contact with a particular social group.

Indices and scales can easily be constructed from multiple indicators that have the same response categories—such

as questions that use the Likert response categories described earlier, which tell us how intensely a person agrees or disagrees with a particular statement. For example, consider several questions from the General Social Survey (GSS) probing the public's views of various governmental institutions. The questions ask about respondents' level of confidence in the "people running these institutions." The response categories for all these questions are as follows:

1. Hardly any
2. Some
3. A great deal

"Hardly any" represents the lowest confidence level, "some" represents an intermediate confidence level, and "a great deal" represents the highest confidence level.

Let's create an index from three GSS questions that measure confidence in the U.S. federal government's executive, legislative, and judicial branches, respectively. Because these three questions compare different dimensions of confidence in the federal government, and responses to one question don't strongly imply a particular response to another question, we will consider this composite measure to be an index rather than a scale.

Because the response categories are exactly the same across the three questions, we can simply sum each respondent's scores across the three questions to create our index. (As we will describe further in **Chapter 14: Quantitative Data Analysis**, things become trickier if your composite items have different response categories, especially if some questions have more categories than others.) **Table 7.5** shows the distribution of summed scores for these questions in the 2021 General Social Survey. Respondents with a sum of 3 have answered all three items with a response of "hardly any" (coded as 1). Therefore, we conclude that they have the lowest confidence in the federal government. Respondents with a sum of 9 have answered all three items with a response of "a great deal." Therefore, we conclude they have the highest possible confidence in the federal government. Respondents with sums from 2 to 8 fall in between these two groups because they have more or less inconsistent responses across the three component items. Overall, our data reveals a rather low level of public confidence in the U.S. federal government, given that more respondents scored near the bottom than near the top.

Table 7.5. Confidence in the U.S. Federal Government: Summed Scores for Respondents' Perceptions of the Three Branches of Government

Confidence in Government Institutions			
		Percent	N
Summed scores scale	3.00 (Lowest)	13.60%	360
	4.00	20.10%	530
	5.00	21.40%	565
	6.00	25.80%	682
	7.00	12.70%	336
	8.00	3.90%	102
	9.00 (Highest)	2.50%	65

Source: GSS 2021.

Note that this index could also be created by *averaging* each respondent's score across the three items rather than *summing* those scores. The index we generated would still obtain a measure of each respondent's general level of confidence in these three governmental institutions, but instead of summed scores, the results would be presented as mean (averaged) scores. **Table 7.6** presents the distribution of those mean scores, ranging from lowest to highest.

Table 7.6. Confidence in the U.S. Federal Government: Mean Scores for Respondents' Perceptions of the Three Branches of Government

Confidence in Government Institutions			
		Percent	N
Mean scores scale	1.00 (Lowest)	13.6%	360
	1.33	20.1%	530
	1.67	21.4%	565
	2.00	25.8%	682
	2.33	12.7%	336
	2.67	3.9%	102
	3.00 (Highest)	2.5%	65

Source: GSS 2021.

The percentages and numbers of cases in Table 7.6 are identical to those in Table 7.5. However, the labels in the table are now means rather than summed scores. If a respondent has a mean score of 1, it means this person has the lowest possible score on all three questions. The only way to get a mean of 1.00 is to have said "hardly any" (coded as 1) for all three questions. On the other end of the scale, the only way to get a mean of 3.00 is to have said "a great deal" for all three questions. The other scores represent mixed responses when averaging the three questions. For example, respondents with a mean score of 2.33 achieved this by answering 3 to one question and 2 to two questions.

Because there is no difference in the percentages and numbers of cases for the two methods—summed scores or mean scores—the researcher could choose either approach. In practice, sociologists may prefer one or the other method of constructing indexes and scales, and you will see examples of both in the literature. The methods section of any academic paper that introduces an index or scale should explain which indicators were included, how they were measured, and how they were aggregated (merged) to create an overall score.

Even if your study involves complex variables, you won't necessarily need to develop your own multiple-item measures. For some variables, you would likely be duplicating work that has already been done by other researchers. That's another reason why it is important to consult the research literature in the early stages of designing your study. For example, if you are operationalizing the concepts of racism or sexism, consider using the Modern Racism or Modern Sexism indices developed by Janet Swim and collaborators (1995). If you are interested in measuring anti-immigrant views, check out Marco Marozzi's index (2016) of perceived threat from immigrants. For attitudes toward transgender individuals, try the 20-item index created by Susan Walch and collaborators (2012). Also remember that scale formats like the Likert scale and Bogardus social distance scale can be applied to a wide range of phenomena—as can the semantic differential scale, Guttman scale, and Thurstone scale.

Using a multiple-indicator measure that's already been developed not only saves you a lot of time and effort, but previously published articles using the measure also will often include information about its quality (specifically, its reliability and validity, the subjects of the next section). You may also find it worthwhile to modify an existing index or scale to match your study's conceptual definitions. Although this can be a good idea in certain situations, note that changing even small parts of a measure can alter its effectiveness. Pilot testing is always recommended whenever you adapt an index or scale for your purposes.

Key Takeaways

1. Sociologists use both indexes and scales to measure complex, multiple-item concepts.
2. Indexes and scales differ from one another in two ways: for a scale, the component items should be conceptually similar to one another (unlike for an index, where they should be conceptually distinct), and they should have some kind of intensity structure.

7.4. Measurement Quality

Learning Objectives

1. Define reliability and validity and discuss how they are related to one another.
2. Describe the three key types of reliability and five key types of validity.
3. Discuss how measurement quality in qualitative research is assessed differently than in quantitative research.

Once we've managed to define our terms and specify the operations for measuring them, how do we know that our measures are any good? Without some assurance of the quality of our measures, we cannot be certain that our findings mean what we think they mean. When social scientists measure concepts, they aim to achieve reliability and validity in their measures. **Reliability** concerns the consistency of our measures. A measure is said to be reliable if it gives the same result when applied repeatedly to the same phenomenon. **Validity** is about a measure's accuracy. A valid measure truthfully reflects the meaning of the concept under study. These two criteria are central to how we evaluate the soundness of measures in quantitative research. As we will discuss later, they can be applied to qualitative research as well, although qualitative sociologists tend to favor other criteria when judging the rigor of their work.

Reliability: A Measure's Consistency



How would you measure whether or not a person is an alcoholic? One possibility is to conduct a survey asking, "Have you ever had a problem with alcohol?" But how do we know whether or not this is a good question? Would a person's response vary according to how much alcohol the person drank on the day just before the survey? If people might give different responses on different days, we have a reliability problem—that is, an issue with the consistency of our measure. *Chris F, via Pexels*

Let's say we want to administer a survey about alcoholism and alcohol intake. We decide to measure alcoholism by asking people the following question: "Have you ever had a problem with alcohol?" If we measure alcoholism in this way, it seems likely that anyone who identifies as an alcoholic would answer this question with a "yes." So, this must be a good way to identify our group of interest, right? Well, maybe. Think about how you or others you know might respond to this question. Would your responses after a wild night out partying differ from what they would have been the day before? Or, suppose you are an infrequent drinker, and the night before you had a single glass of wine—which gave you a headache on the day you're taking the survey. Would that change how you might answer the question? In each of these cases, it's possible that the same person would respond differently *at different points in time*. This suggests that our measure of alcoholism has a reliability problem.

Memory often creates issues with a measure's reliability. If we ask research participants to recall some aspect of their past behavior, we should try to make the recollection process as simple and straightforward for them as possible. For example, if we ask respondents how much wine, beer, and liquor they've consumed each day over the course of the past

three months, how likely are we to get accurate responses? Unless a person diligently keeps a journal documenting their intake, they will likely give us answers to that question that are all over the map—maybe accurate, but maybe wildly inaccurate. In contrast, if we ask the respondent how many drinks of any kind they have consumed in the past week, we should consistently get more accurate responses. People will be better able to remember what happened over a week's time. As a result, the second question is a more reliable measure.

Reliability can be an issue even when we're not reliant on others to accurately report their behaviors. Perhaps you are interested in observing how alcohol intake influences interactions in public locations. You may decide to conduct observations at a local bar, noting how many drinks patrons consume and how their behavior changes as their intake changes. But what if you go to the restroom and miss the three shots of tequila that the person next to you downs during the brief period you are away? The reliability of your measure of alcohol intake—counting the numbers of drinks that patrons consume—depends on your ability to actually witness every time that a patron consumes a drink. If you are unlikely to be able to observe every such instance, then perhaps your procedure for measuring this concept is not reliable.

When designing a study, you should keep in mind three key types of reliability. First, multiple applications of a measure over time should yield consistent results (assuming that the phenomenon being measured has not changed). For example, if you took standardized tests to measure your college readiness on multiple occasions in your senior year of high school, your scores should be relatively close from test to test. This is known as **test-retest reliability**. In the same way, if a person is clinically depressed, a depression scale should give results today that are similar (although not necessarily identical) to what it gives two days from now. In both these cases, we do not have reason to believe that the phenomenon being measured—scholastic aptitude or clinical depression, respectively—has changed in any substantial way, so the results across those applications should be similar.

If your study involves observing people's behaviors—for example, watching sessions of mothers playing with infants—you may also need to assess **inter-rater reliability**. Inter-rater reliability is the degree to which different observers agree on what happened. If you and another researcher both watched a parent interacting with their child, would the two of you evaluate the parent's engagement in a similar way? If your scores varied wildly, we might suspect that how you are measuring parental interactions is unreliable. Perhaps you identified certain behaviors as signs of an engaged parent that your fellow researcher did not. Or perhaps you are a keenly observant person, and you caught moments that the other researcher missed—such as when the infant offered an object to the mother, and the mother dismissed it. A similar issue arises if you are reading through interview transcripts or conducting a content analysis. Your interpretation of what is being said or what themes arise in your data may differ dramatically from another person's version.

A research team can reduce these problems of inter-rater reliability by establishing clear and well-enforced standards for conducting observations and categorizing the data. They may write a formal document with guidance and examples regarding how to evaluate certain situations or outcomes. When making sense of a particular piece of data, such as an interview transcript, they may have two or more people review the same data to see if there are major differences in the conclusions they individually draw; if so, more training, documentation, and discussion could bring people's perspectives into closer alignment.

With such safeguards, the scores of different observers should be consistent, although perhaps not perfectly identical. If a single person is doing all the observations for a study, however, there's no way of checking whether that person's observations would diverge significantly from another person's. If possible, you should measure inter-rater reliability during pilot testing—enlisting colleagues to help you as needed—so that you can put procedures in place to narrow any such gaps. For published work, research teams are often required to include a measure of inter-rater reliability in their write-up to ensure that they are upfront about any problems in this area. Even solo researchers must acknowledge potential problems and defend their analysis from charges that it is fundamentally unreliable.

When crafting multiple-item measures, **internal consistency** is a third important thing to consider. The scores on each

question of an index or scale should be correlated with each other, as they all measure aspects of the same overall concept. Let's say we created an index that measures support for feminist issues. It includes questions about whether the person would vote for a woman for president and whether they support paid family leave. If we ask a group of people these questions, we would expect to find a strong correlation between the scores for each question. A person who would vote for a woman for president would probably also express support for paid family leave, for instance. The two questions don't capture the exact same thing, but if both are feminist issues, they should be related in terms of how people answer them. This should be the case for all the questions in the index or scale if it is internally consistent. (As we will discuss further in **Chapter 14: Quantitative Data Analysis**, a specific statistical test known as Cronbach's alpha—written as α—provides a way to measure how closely answers to each question in an index are related to those for the other questions; you will frequently see this figure reported in papers that use indexes or scales.)

Ideally, you should evaluate these three types of reliability—test-retest, inter-rater, and internal consistency—both in your study's pilot-testing process and after your data collection is done. That said, the design of your study will affect your ability to do so. For example, if you are conducting a longitudinal study that involves collecting data from the same participants at multiple times, you can evaluate test-retest reliability. You won't be able to do so, however, in a cross-sectional study after the data collection is complete.

Validity: A Measure's Accuracy



If a person goes to the local gym regularly but just hangs out in the hot tub, would we say they are engaging in healthy behavior? Probably not—which might make us wonder if just counting the number of times a person goes to the gym every month is really a good measure of whether they lead a healthy lifestyle. With this critique in mind, we'd say the measure lacks face validity: on the face of it, it seems that monthly gym visits might not be the best indicator of engaging in a healthy lifestyle. This does not mean the measure is totally invalid—remember that sometimes we need to make do with less-than-ideal procedures, given practical constraints in our research—but we'd probably want, at the very least, to measure other things as well so that we avoid including so many hot tub enthusiasts in the group that we determine to be super-healthy. *Ron Lach, via Pexels*

While reliability is about whether a measurement procedure gives consistent results, validity (also known as *construct validity*) is about whether the procedure actually measures the concept it is intended to measure. Put a bit differently, valid measures accurately get at the meaning of our concepts. Think back to the measures of alcoholism we considered earlier. We initially considered asking research participants the following question: "Have you ever had a problem with alcohol?" We realized that this might not be the most reliable way of measuring alcoholism because the same person's response might vary dramatically depending on how they are feeling that day. But this measure of alcoholism is not particularly valid, either. For one thing, it's unclear what we mean by "a problem" with alcohol. Some people might think they have "a problem" after experiencing a single regrettable or embarrassing moment after consuming too much. For others, the threshold for "a problem" might be different: perhaps a person has had numerous embarrassing drunken moments but still gets out of bed for work every day, so they don't perceive themselves as having a problem.

Because what each respondent considers to be problematic might be very different, our measure of alcoholism isn't likely to yield any useful or meaningful results if our aim is to objectively understand, say, how many of our research participants are alcoholics. (Granted, the definition of alcoholism is partly subjective and open to debate, but this particular question remains a poor measure of whether the person is an alcoholic based on any common understanding of the term.) Note that the problem isn't the fact that the *same* person will respond to the question differently at different times—a problem of reliability—but that *different* people will interpret the question in different ways—a problem of validity.

There are different ways of evaluating whether a given measure is valid. When we were arguing that the question "Do you have a problem with alcohol?" is a flawed measure of alcoholism, we were using logical reasoning to assess its **face validity**. Face validity asks whether it is plausible that the question measures what it intends to measure. On the face of it, does the question pass muster? Logically speaking, can it measure what we want it to measure? Sometimes determining face validity is easy. For example, a question about a person's place of residence would clearly lack face validity if we were using it as a measure of their political views; even if people's politics often correlate strongly with where they live, we should really be asking them directly about their political opinions. Other times, however, face validity can be more difficult to assess. Let's consider another example.

Suppose we're interested in learning about a person's dedication to "healthy lifestyle" practices. Most of us would probably agree that engaging in regular exercise is a sign of dedication to a healthy lifestyle, so we could measure this concept by counting the number of times per week that a person visits their local gym. One of the first things we'd have to consider is whether that operationalization has face validity. Does our operationalized variable measure the concept that it is intended to measure on the face of it? In the case of counting the number of times per week visiting a local gym, perhaps they want to use their tanning beds, flirt with potential dates, or sit for hours in the sauna. These activities, while potentially relaxing, are probably not the sort of "exercise" we had in mind as one dimension of a healthy lifestyle, and therefore, recording the number of times a person visits the gym may not be the most valid way to measure our "healthy lifestyle" concept on the face of it. Here, the problem isn't that different people will interpret the same question differently, but rather that the question doesn't quite measure what we want it to measure.

One method for assessing face validity is to pilot test measures with research participants who are like those who will be in the final study group. For example, if you were conducting a survey of transgender individuals about their access to appropriate healthcare services, you could attend transgender-related events and solicit a small number of volunteers to test your questionnaire. Following a common testing method, you could first have the pilot-testers answer the questions, and then ask them to identify any questions that were puzzling, ambiguous, or confusing. You could have the pilot-testers explain why they felt this way, thereby gaining some insight into how the survey questions might have faltered in measuring your study's underlying concepts. For example, let's say one of your questions was "Have you ever experienced discrimination by a doctor or other healthcare provider due to your transgender status or gender expression?" In your initial draft of that question, you only gave respondents the response categories "Yes" and "No." In your pilot testing, however, your respondents mentioned that they did sometimes feel that providers discriminated against them, but they were not sure if it had anything to do with "transgender status or gender expression." Based on this feedback, you ultimately decide to provide an "unsure" response option, ensuring face validity and thereby enhancing the overall validity of your measure.

To illustrate another type of measurement validity, let's examine the sorts of operationalization decisions that might relate to our hypothetical study of transgender access to health care. For this study, we are interested in how transgender people learn about the healthcare services available to them. We decide to operationalize this concept—"sources of information about transgender health care"—by asking our respondents a question: "Have you received any information about transgender health care from your doctor?" If we think through the various ways that transgender people might learn about their healthcare options, however, we would soon realize that this single question

is not enough: talking to their doctor is not the only way that our respondents could receive this sort of information. As a result, our single indicator does not cover all of the possible meanings of our concept.

Content validity assesses our measurement in this way: does our chosen measure get at all the important meanings or dimensions of our concept? "Sources of information about transgender health care" is a multidimensional concept that really needs an index or scale to measure it comprehensively. In addition to asking our respondents about whether they received any information from doctors, we need to ask about other possible sources, such as friends, transgender support groups, and the internet.

Always consider whether aspects of your concept of interest aren't included in the particular measure you're using to examine it. This is important regardless of whether you create the measure yourself or borrow it from previous research—in fact, one way that social scientists can contribute to the literature is by adapting past measures to current times (such as how social scientists have changed the measures they use to identify and count "families" in a given society, as we discussed in an earlier chapter). If you are unsure whether you are measuring all the relevant dimensions of a concept, you may need to do some more digging first—reading the relevant literature or conducting pilot interviews or focus groups with people who know that concept well (such as the transgender people in the previous example). Sometimes researchers will even convene panels of experts on the subject to reach some sort of consensus about what the different dimensions of a given phenomenon are and how best to measure them.

A third crucial kind of validity is **predictive validity**. Let's say we have found an existing index that seems to measure the concept of "dedication to healthy living" in a rich and multifaceted way. To make sure our measure is truly valid, we don't want to stop there—we should test how well it predicts relevant outcomes. For example, based on how our respondents score on our index of healthy living, we should be able to make good predictions of how well they will do on various health evaluations—say, how favorably the blood panel tests they take during their annual physicals turn out. If our measure is effective in this way, we say it has predictive validity: it predicts future phenomena that it should be able to predict, considering the underlying concept it is measuring.

Smart researchers are always on the lookout for ways to test the predictive validity of any measure that they or others choose to use. Sometimes, testing those predictions can become a matter of fierce debate. We already mentioned in the sidebar ***The Bell Curve* and Its Critics** how social scientists have challenged the utility of the IQ measure, and a similar controversy surrounds the SAT and similar standardized exams used as the basis for admissions decisions by colleges and universities. If SAT scores are a truly valid measure of "scholastic aptitude," they should be able to predict student success in college—the higher a person's score, the greater the likelihood they will do well in their college classes. Research on this question finds that high SAT scores are associated with high grades in college, but some social scientists argue that the SAT is nonetheless a flawed predictor that largely reflects the fact that kids from more advantaged backgrounds do well both on the SAT *and* in their college classes.[1]

Another good way to test a measure's validity is to look for an existing measure of the same (or at least a similar) concept. You should use both measures when you collect your data and then analyze the results you get from each. If their scores are correlated, then you have reason to believe that both are measuring the same concept. This is called **convergent validity**: the two measures converge on a similar result. For example, if we wanted to evaluate our index of healthy living,

1. Research that tries to control for this confounding variable of socioeconomic status, such as work by the economist Jesse Rothstein (2004), finds that the SAT predicts little once those background characteristics are accounted for. The research also generally finds that high school grades are more effective than SAT scores in predicting college performance, although there is considerable debate over whether considering SAT scores alongside high school grades substantially improves that prediction; testing companies argue that using grades alone might worsen existing inequalities because grade inflation is more rampant in affluent areas (Kurlaender and Dykeman 2019; Watanabe 2019).

we might turn to the Healthy Lifestyle Behaviors Scale (HLBS) (Walker, Sechrist, and Pender 1987). If we gave someone our healthy living index and the HLBS at the same time, their scores should be pretty similar.

Testing for **discriminant validity** uses the exact opposite approach. In this case, you compare your measure to a measure that, logically speaking, should be entirely unrelated. For example, our respondents' scores on our healthy living index should not be strongly related to their scores on, say, a scale measuring the intensity of their political opinions. If they *are* related, then we have reason to question whether we're really measuring what we intend to be measuring. Our assessment can be sharper if we compare our measure's discriminant validity with its convergent validity. For example, we might reasonably expect not only that a measure of a person's self-esteem will converge with a measure of their confidence in social interactions but also that the association we uncover will be stronger than the one between a person's self-esteem and some clearly less relevant measure, such as their support of authoritarianism.

Thinking about Reliability and Validity in Research

We've touched upon many aspects of reliability and validity in the preceding sections, and by now you've probably mixed them up hopelessly in your heads. We've got you covered: **Figure 7.7** illustrates a classic way of visualizing the differences between these two ideas.[2] Across these three targets, think of the placement of the shots as a sociologist's attempts to measure a particular concept. *The reliability of the measure—that it generates the same results under the same conditions—is represented by the consistency of the shots on the target*: the tighter the pattern of shots, the higher the measure's reliability. In the left-hand target, we see a wildly inconsistent pattern of shots—in other words, low reliability. In *both* the middle and right-hand targets, however, the shots are tightly clustered in a single area, signaling high reliability.

Now let's consider the three targets in terms of validity. *The closer the shots are to the bullseye—which represents the actual meaning of the concept—the higher their validity.* In the left-hand target, the sociologist's measure has a low level of validity: few of the shots are near the bullseye, meaning that the measurements do not accurately reflect the underlying concept. In the right-hand target, all the shots are on the bullseye. The measure is valid. The sociologist is measuring the right thing. Now, the middle target is particularly interesting. The measure being depicted clearly has high reliability, as all the shots are tightly clustered. Yet this measure would be unwise to use because its validity is low—it's not measuring what the sociologist intends to measure. That speaks to an essential truth about research: there's little use in being consistent if you are consistently wrong.

You also may be wondering why the diagram has only three of the four possible high-low combinations? Why didn't we include a target illustrating low reliability and high validity? We left it out to make a point: it's harder to hit a bullseye on a target (i.e., obtain highly valid measurements) when your shots are scattered (your measure has low reliability). A research project that gathers only unreliable measurements is unlikely to accurately capture its underlying concepts of interest and to generate valid results.

2. See, for example, Earl Babbie's *Practice of Social Research* textbooks (2010:153).

A final word of advice: as you read more sociological studies, you will find that there are many more ways to evaluate the reliability of measures rather than their validity. This is because assessing the consistency of a measure is generally more straightforward—as you can see in the simple procedures we just described. In contrast, social scientists engage in endless debates about whether a particular indicator or index is truly the best way of capturing the essence of a particular phenomenon. Just think about our earlier discussion of the numerous methods that scholars and governments use to measure poverty. This particular methodological debate is especially partisan and polarized, but you should be aware of how complicated an undertaking it is to arrive at a valid measure of many social concepts—even ones that aren't as obviously political. You will have to be exceedingly careful in how you approach the steps of conceptualization, operationalization, and measurement we illustrated in **Figure 7.1** to ensure that your study ultimately generates valid results. Among other things, you should pay attention to the process represented by the arrow curving back upward: always go back to the key concepts of your study and ensure that your measures truly reflect them.

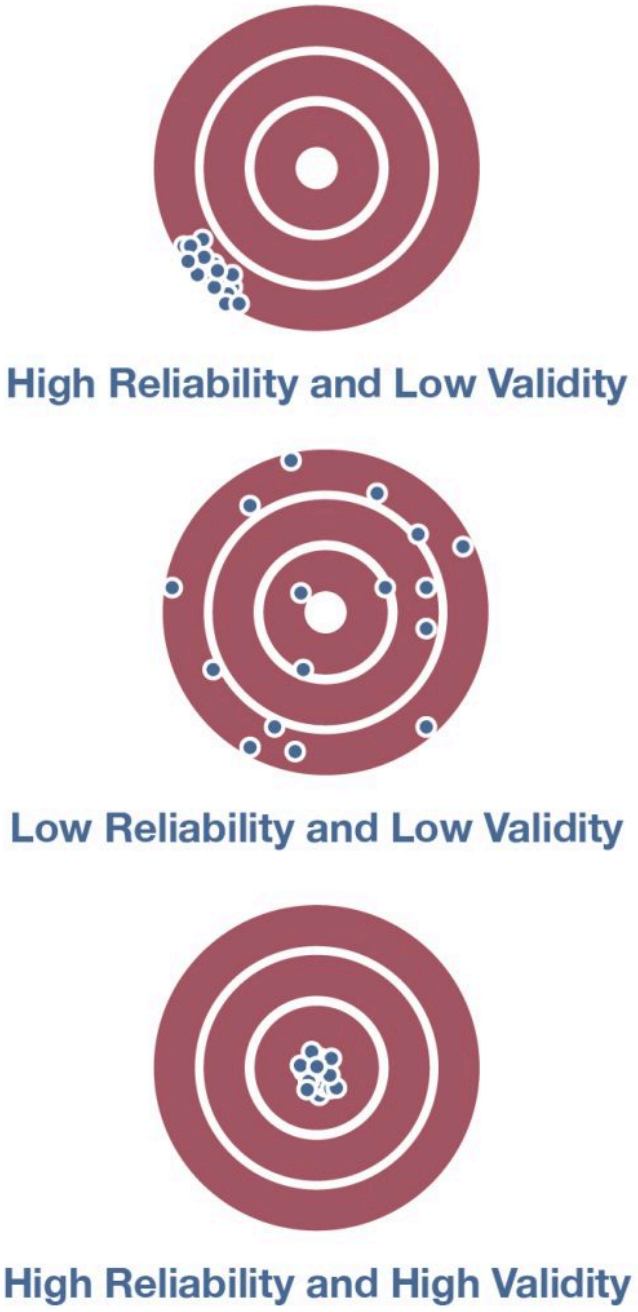


Figure 7.7. Comparing Reliability and Validity. A measurement procedure is reliable if it gives us the same results under the same conditions, and it is valid if it actually measures the concept we intend to measure. The patterns of shots on the three targets provide a visual way of thinking about the reliability and validity of our measures. How close the shots are to each other tells us how reliable that particular measure is: a tight cluster means high reliability, and a scattered pattern means low reliability. How close the shots are to the target's bullseye tells us how valid our measure is: near the center means high validity, and far from the center means low validity.

Measurement Error: Random and Systematic

As we have described, good measures demonstrate validity and reliability. Nevertheless, when we actually take measurements out in the real world, they always have some degree of error. **Measurement error** is the difference between the values of a variable and the true—but unobserved—values of that same variable. For example, measurement error occurs when a person tells the interviewer her age is 46 when it's actually 43, or when the U.S. Census Bureau counts the population of homeless people as lower than it actually is.

Measurement error can take two forms: *random* and *systematic*. **Random error** is present in any measurement. Even with the most precise instruments, our measurements are always a little bit off—never perfectly describing what really exists. That's to be expected. If you've ever stepped on a bathroom scale twice and gotten two slightly different results—maybe a difference of a tenth of a pound—then you've experienced random error.

Maybe you were standing slightly differently or had a fraction of your foot off of the scale the first time. If you were to take enough measures of your weight on the same scale, you'd be able to figure out (or at least get closer to) your "true" weight.

A similar kind of random error plagues our measurements in the social sciences. For example, if you gave someone a scale measuring depression on a day when they lost their job, they would likely score higher on it than on a day when they received a promotion. Even if the person was truly suffering from clinical depression, the random occurrences of that person's life would alter our measures of that depression—much like a stray foot on the bathroom scale. For this reason, social scientists are taught to speak with humility about our measures. With all the factors at play when dealing with complex human beings, it's exceedingly hard to reduce the random error that plagues social scientific measures. With enough rigor in our approach, we can be reasonably confident that what we have found is true, but we must always acknowledge that our measures are only an approximation of reality.

Although we should be resigned to some degree of random error in our measurements, **systematic error** is more worrisome. This issue arises when a measure consistently produces incorrect data, typically in one direction—making our measurements consistently different (say, too high or too low) compared with the true value. Imagine you surveyed students at your local college or university and asked them to give their height, but you didn't provide any response categories for heights taller than six feet, six inches. Some of the taller members of the basketball team might not be measured accurately. The average height you calculate based on this data would not match the real average height of the student body—it would be lower by a certain amount, because the tallest people would have been miscategorized as shorter than they actually are.

Researchers can also create systematic errors in their measures when they use question wording that causes respondents to think one answer choice is preferable to another. For example, if an interviewer were to ask you, "Do you think global warming is caused by human activity?" you would probably feel comfortable answering honestly. But suppose the interviewer asked, "Do you agree with 99 percent of scientists that global warming is caused by human activity?" Would you feel comfortable saying no, if that's what you honestly felt? Maybe not. That is an example of a **leading question**, a question with wording that encourages a respondent to answer in a certain way. (We discuss leading questions and other problems in question wording in greater detail in **Chapter 13: Surveys**.) Note that a poorly worded question sometimes creates confusion, but it does not prod respondents to answer *in a particular way*. In this case, the error is random, not systematic. We should be concerned that our measure is unreliable—people may answer very differently depending on how much they understood the question—but we don't have reason to believe that the answers to our question are skewed in a particular direction. That's not the case, unfortunately, for the sorts of leading questions we just mentioned, or for questions that suffer from other forms of **bias**—that is, factors that cause our study to give us a picture of social reality different from the true state of affairs.



The way we ask a question may encourage certain responses, which can systematically skew our results toward those responses. For example, if we ask people if they wash their hands after using the bathroom, they will want to present themselves in a socially desirable light, and many will answer "Yes" even when they are actually slackers in this regard. (This kind of systematic error is called *social desirability bias*.) In this case, we are likely to get more "Yes" answers than we would have if people answered truthfully. *Ketut Subiyanto, via Pexels*

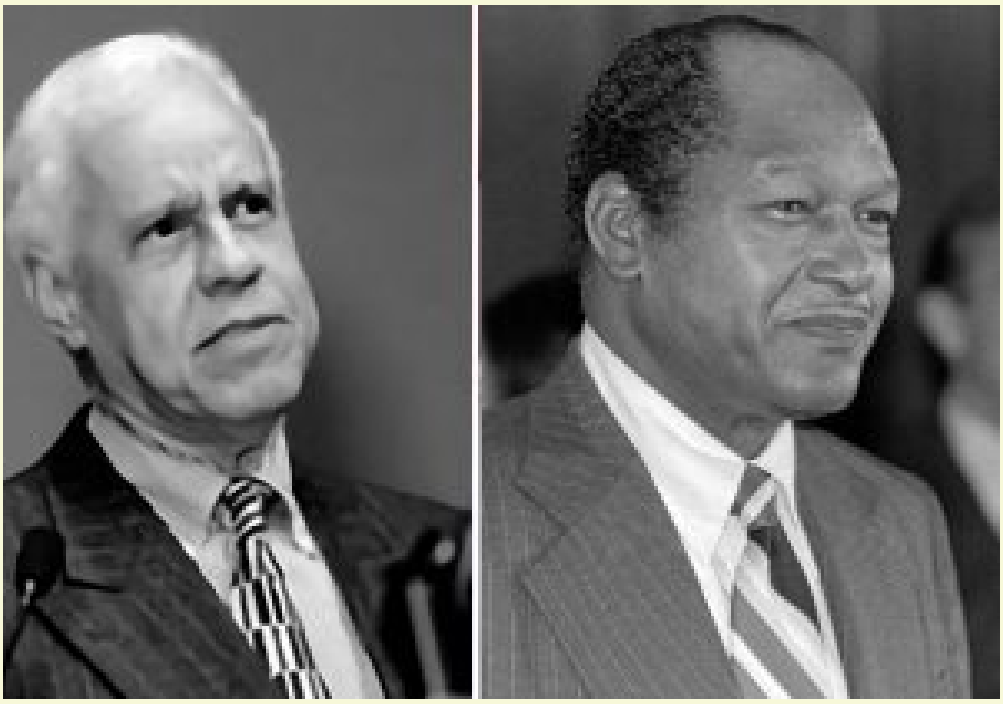
Imagine you were conducting a survey that asked people if they washed their hands after using the bathroom. Would you expect your respondents to be perfectly honest? Their answers might say more about what they would like others to think they do, rather than what they actually do. When participants in a research study answer or act in particular ways to present themselves to the researcher in a more positive light, we call this kind of bias **social desirability bias**. Generally speaking, people in any study will want to seem tolerant, open-minded, and intelligent, even if their true feelings are biased, closed-minded, and simple. This is a common concern in political polling. Pollsters are constantly worried about whether respondents are being honest about supporting a controversial candidate—or not supporting a third-party candidate, a person of color, or a female candidate. (See the sidebar **Social Desirability in Political Polling**.)

A related form of measurement error is called **acquiescence bias**, also known as "yea-saying." It occurs when people say "yes" to whatever the researcher asks, even when doing so contradicts previous answers. For example, a person might say yes to both "I am a confident leader in group discussions" and "I feel anxious interacting in group discussions." Those two responses are unlikely to both be true for the same person. Why would someone do this? One reason is social desirability: people want to be agreeable and nice to the researcher asking them questions. They might be trying to "save face" for themselves or the person asking the questions. Or, they might simply want to ignore their contradictory feelings when responding to each question. Regardless of the reason, the results of your measure in this particular situation will not match what the person truly feels. We could imagine a study in which more "yes" answers would support a particular finding, which would mean a large amount of systematic error that biases the study in the direction of that finding.

Usually, systematic errors will result in responses that are incorrect in one direction or another. For example, social desirability bias usually means more people will *say* they will vote for a third party in an election than actually do. Systematic errors of any kind can be a problem, but they are especially problematic when they skew our results in a particular direction—meaning that we grievously overstate or understate the prevalence of a phenomenon or the severity of a problem. Fortunately, we can take practical steps to reduce systematic errors. As we will discuss later in **Chapter 13: Surveys**, we can word questions in such a way to make it more likely that people will speak honestly. We can mix up our affirmative and negative questions so that acquiescence bias doesn't push our study's findings in one direction.

We can never eliminate random error, but by definition, this other type of measurement error is unpredictable and does not result in scores that are consistently higher or lower on a given measure. You can think of random error as statistical noise that you can expect to average out across participants. Like static in your headphones, random error may make it harder to make out certain things, but it won't skew your statistics or your overall sense of the patterns in the data—which is a good thing.

Social Desirability in Political Polling: The Wilder Effect and Shy Trump Voters



Polling before the 1989 election for Virginia governor may have been subject to social desirability bias. Polls had L. Douglas Wilder (*left*) leading by a much larger margin than he ended up winning by. Some observers questioned whether voters had been reluctant to admit to pollsters that they would not actually vote for an African American candidate. A similar claim was made about the 1982 California gubernatorial race. Tom Bradley led in preelection polling, but ended up losing the election. *NIH, via Wikimedia Commons / UCLA Library, via Wikimedia Commons*

The 1989 election to decide Virginia's governor pitted Democratic candidate L. Douglas Wilder against Republican candidate Marshall Coleman. Wilder was running to be the first African American elected governor of a U.S. state. Predictions from statewide polls during the campaign indicated that Wilder was heavily favored

to win election. “Two weeks out from the race and maybe shortly before that, there were people proclaiming victory by saying that I was up by double digits,” Wilder recalled in an NPR interview decades later (Martin 2008). “Several polls said there was no way I could lose.” Wilder did indeed beat his white challenger—but he won by less than 1 percent of the vote. It turned out that both preelection political polls and election-day exit polls had wildly overestimated Wilder’s actual support among voters. The race ended up being so close that Wilder had to wait for a recount to confirm his win.

This was not the first time that political polls had blatantly overestimated an African American candidate’s support among the U.S. electorate. Tom Bradley, a former Los Angeles mayor, ran for governor of California in 1982. A poll taken about a month before the election had Bradley up by 22 points. Another poll conducted close to the election day had him with a 10-point lead. Bradley went on to lose in the closest election in California history (Payne 2010). After Bradley’s unexpected defeat, social scientists started talking about the “Bradley effect”—the notion that voters didn’t want to admit to pollsters that they weren’t actually going to vote for a nonwhite candidate. After Wilder’s near-loss, the term “Wilder effect” also came into circulation to describe this type of alleged polling bias.

As sociologists who conduct surveys know well, respondents will try to answer sensitive questions in a way that will present them in a favorable light. This so-called social desirability bias led to the flawed polling in California and Virginia, proponents of the Bradley/Wilder effect argued. The two men were running for political office during a period when voters were becoming increasingly reluctant to admit that they would not vote for a nonwhite candidate. Social desirability bias caused some respondents to tell pollsters they favored the African American candidate in these two races when they did not actually plan on voting for him on election day. The Bradley/Wilder effect has since become a popular theory in political circles. It remains controversial among social scientists, however, given the challenge of rigorously testing it. Muddying matters even further, the political scientist Dan Hopkins (2009) analyzed U.S. Senate and gubernatorial races from 1989 to 2006 and concluded that the Bradley/Wilder effect was present in political polls before 1996 but—for unclear reasons—vanished in later elections.



Prospective voters in the 2016 presidential election between Hillary Clinton and Donald Trump may have been reluctant to admit they were planning to vote for Trump, a candidate particularly noteworthy for generating both intense enthusiasm and intense dislike. Some researchers have argued that the polls that year undercounted "shy Trump voters" who, because of social desirability bias, wouldn't tell pollsters how they would actually vote—although empirical support for this claim has been mixed. *Michael Anthony, via Pexels*

The surprising outcome of the 2016 presidential election—which, according to most preelection surveys, the Democratic candidate Hillary Clinton had been favored to win—generated new allegations that social desirability bias was once again skewing the polls. This time, political observers argued that the highly polarized attitudes about Republican Donald Trump's campaign for president had made his supporters reluctant to admit they were planning to vote for him—or reluctant even to respond to pollsters from establishment organizations. These votes for Trump would not have been picked up by the preelection polls, which could explain why turnout for Trump ended up being much higher than predicted. As with previous electoral mispredictions, a compelling theory emerged after the 2016 election to explain the polling industry's terrible, horrible, no good, very bad day. Despite all the airtime that this notion of "shy Trump voters" has sucked up on cable news in the years since, empirical support for it remains mixed (Coppock 2017; Enns, Lagodny, and Schuldt 2017).

If the size of these so-called Bradley/Wilder/Trump effects may be exaggerated, social scientists do worry a great deal about social desirability bias in their surveys—political polls or otherwise. There are some straightforward ways to reduce it (as we describe further in **Chapter 13: Surveys**), and more involved

approaches can be undertaken when surveys need to ask about particularly sensitive topics. For example, survey methodologists have created "social desirability scales" to identify which respondents are more likely to offer up socially acceptable answers. Probably the most well-known is the Marlowe and Crowne Social Desirability scale (MCSD), developed in 1960 and still in use today, which contains four items in one condensed version of the scale (Haghighat 2007):

1. Would you smile at people every time you meet them?
2. Do you always practice what you preach to people?
3. If you say to people that you will do something, do you always keep your promise no matter how inconvenient it may be?
4. Would you ever lie to people?

The answers to these questions are used to generate an index score. The higher the score, the more likely respondents are to give socially desirable answers. The scale can be used to adjust the results of surveys—including preelection polls—to account for the fact that social desirability may be skewing expressed opinions upward or downward.

Social desirability bias is by no means the only reason that political polling may be wildly off in its electoral predictions. Far more important, as we discussed in **Chapter 6: Sampling**, is the selection bias that affects who actually is being represented in a poll's sample. Furthermore, preelection polls by definition happen before election day, and political events can dramatically change a candidate's level of support within a brief span of time—as arguably happened in the 2016 presidential election, with Clinton plagued by damaging news stories in the final days of her campaign.

Talking two decades later about his close gubernatorial race, Wilder himself speculated that the so-called Wilder effect wasn't so much about voters reluctant to tell the truth but rather polls drawing too much from Democratic-leaning areas (Martin 2008). Remember that political polls are surveys, and surveys can fall short in their measurement and sampling for countless reasons.

Measurement Quality in Qualitative Research

In qualitative research, the standards for measurement quality differ in important ways from those for quantitative research. Data collection in quantitative research involves a level of objectivity and impartiality that qualitative research cannot match. In a straightforward fashion, quantitative researchers will choose a measure, apply it, and then analyze the results. As we have noted, however, qualitative researchers typically don't ask the exact same questions of every respondent or conduct the exact same kind of observations in every space.

Rather than relying on fixed measurement instruments, qualitative researchers should see *themselves*, in effect, as their measurement instruments. What does this mean, in practice? For one thing, qualitative researchers must make full use of their rapport with research participants and their immersion in particular social spaces. This closeness to the data allows them, in turn, to present a rich picture of an individual or community's lived experiences and subjective understandings and to describe in rich detail the social processes they observed. In the course of their fieldwork, qualitative researchers build connections between the different ideas that participants discuss. They ask follow-up questions or conduct additional observations to delve more deeply into how exactly a concept is understood or experienced. They then draft an analysis that accurately reflects the depth and complexity of what respondents have said or what they have witnessed.

This is a challenging set of tasks to do well. It requires balancing competing priorities. For instance, qualitative

researchers always need to acknowledge their personal biases—either from their own experience or their existing knowledge of the topic—given that the very way they are gathering data makes it difficult to be truly objective and methodical (we will discuss the idea of **positionality** further in **Chapter 9: Ethnography**). At the same time, they have to be careful about accepting their respondents' statements or perspectives uncritically. Qualitative sociologists carry a toolkit of useful theories and methodological skills into the field, but they also need to be open to the ways that the data they gather can contest or complicate the understandings they originally brought with them. In fact, this is often the analytical approach that qualitative researchers take in their fieldwork: not trying to prove or disprove a hypothesis, but rather looking for unexpected connections, complexities, or wrinkles in whatever theoretical perspective they're using to view the world.

For these reasons, good qualitative research hinges not so much on being objective—there is always some subjectivity in qualitative analysis—but more on the degree of rigor that the individual researcher brings to bear in their fieldwork and analysis. The "rigor" we have in mind is not about validity and reliability per se. Although it is possible to critique qualitative research in these terms, many researchers use different language to describe the soundness of qualitative work. Two criteria often used are *authenticity* and *trustworthiness* (Rodwell 1998).

Authenticity refers to the degree to which qualitative researchers capture the multiple perspectives and values of participants in their study. While there are multiple dimensions of authenticity, the most important one for novice researchers to understand is **fairness**. Are you considering all the relevant viewpoints on a particular topic that exist, or have you narrowed them down to the ones you, the researcher, prefer? Are you favoring a particular explanation or particular understanding over others, without good reason? To ensure fairness, qualitative sociologists spend considerable time and effort reflecting on the biases and blind spots of their work—which is part of the reason that good qualitative work takes so long to publish. Frequently, qualitative sociologists also take steps to equalize the power dynamics in their relationships with research participants. For instance, they may try to focus their research on questions that the community being studied wants answered. They may involve participants in the data analysis process, working closely with them to arrive at a consensus on the final interpretation of the data.

Trustworthiness refers to the overall truthfulness and usefulness of the results of the research study. We can think of trustworthiness in three dimensions (*credibility*, *dependability*, and *confirmability*), which resemble the criteria of validity and reliability used to evaluate quantitative studies, though with important differences in emphasis.

Credibility refers to the degree to which the results are accurate and viewed as important and believable. Credibility is akin to validity, because it speaks to how closely the study reflects the reality of the people being studied. To ensure the credibility of their work, qualitative researchers sometimes run it by their research participants before publication. The idea is that the published work should make sense to the people who actually live or work in the spaces being studied—if it doesn't jibe with their lived experiences, it's not a credible account. Of course, it is also true that people immersed in a particular social setting may not be able to see what an outsider sees about it. More generally, seeking to make one's work credible to the people being studied can have trade-offs, too, which is a chief reason that most journalists do not follow a similar practice in their reporting. (In their view, showing a story to sources before publication can create pressures to change their reporting and taints the objectivity they strive for in their work.) Nevertheless, qualitative sociologists should think long and hard about whether their study's findings faithfully capture what happens in the social worlds being studied, and whether it is ultimately credible—to the study's participants and its readers alike.

The criterion of **dependability** is similar, but not equivalent, to reliability. Much qualitative work is anything but reliable in the narrow way we previously described—ask the same question, get similar answers. As we have noted, qualitative research questions, hypotheses, and data collection processes routinely change during fieldwork. Qualitative research assumes that new concepts and connections will emerge from the data collection, and because of this process of exploration, respondents never get the exact same questions each time they are interviewed. Indeed, because qualitative research emphasizes the importance of context, it goes against the strength of the method to try to ask a particular question under the exact same conditions for all respondents. Can qualitative researchers achieve reliable results

under such conditions? Not really. Instead, they aim for dependability. Their work can be considered dependable if they have followed the systematic research procedures and best practices for interviewing and observation that qualitative sociologists have developed over the years (we describe these methods in detail in **Chapter 9: Ethnography** and **Chapter 10: In-Depth Interviews**). Their work can also be considered dependable if they adequately documented their methodology and any changes made to it over the course of their fieldwork and analysis.



Sociologist Matthew Desmond's book, *Evicted*, covered controversial topics. To help ensure data quality, Desmond took the unusual step of hiring a professional fact-checker to evaluate his materials. This extra step increased the *credibility* of the text (the degree to which a study's results are accurate and believable) and also gave readers more reason to trust its *confirmability* (the degree to which any reported findings can be traced back to specific interviews, observations, and other data). *U.S. Library of Congress, via Wikimedia Commons*

Finally, the criterion of **confirmability** refers to the degree to which the results reported are actually grounded in the data obtained from participants. Although it is possible that another researcher could view the same data and come up with a different analysis, confirmability ensures that a researcher's findings have a firm basis in what participants said. Another researcher should be able to read the results of your study and trace each point made back to your data—something specific that one or more respondents shared in interviews, or something specific you observed. (This process of review is called an **audit**.) Any patterns observed in your data should not be exaggerated or otherwise mischaracterized in your research product.

In recent years, some public sociologists have followed journalists in hiring independent fact-checkers not just to conduct audits—ensuring that the write-up matches the collected data and its patterns—but also to confirm that every fact in their published work is accurate. For example, for his Pulitzer Prize–winning book *Evicted* (Desmond 2016), the sociologist Matthew Desmond took the extra step of paying for professional fact-checking of his research. Few sociologists can afford to do this, but higher levels of vetting would seem to be wise for qualitative work with any potential to spark intense public scrutiny—especially given the risks of the sorts of backlash described in **Chapter 8: Ethics**.

It remains to be seen whether intensive audits and fact-checking will become the future of qualitative sociology, but in the meantime, even researchers on a shoestring budget can benefit from sharing their work with colleagues—peers and professors—before publishing it. As you might expect, it's difficult to view your own research without bias, so another set of eyes is often helpful. Careful reviewers should be able to raise helpful questions about the confirmability, credibility, and dependability of your work. Specifically, they can tell you if your work sounds plausible to them, given what exactly you are claiming and what evidence you provide to support those claims. They can point out any key flaws or omissions and push you to justify your methodological decisions more clearly and convincingly. They can examine your fieldnotes, documentation, and results to determine whether you followed proper research procedures. Obviously, any data you share with others will have to be properly anonymized to make sure you don't violate any confidentiality agreements with your respondents, so sharing the final research product is often much easier than opening up your field notes or interview transcripts for scrutiny. That said, even just contemplating that level of transparency will often make you, the qualitative researcher, more judicious about the sorts of arguments you make based on the data you have.

Key Takeaways

1. Reliability is about consistency of measurement, and validity is about accuracy of measurement.
2. The three key types of reliability are test-retest, internal consistency, and inter-rater reliability.
3. Five key types of validity are face, content, predictive, convergent, and discriminant validity.
4. Qualitative research uses different criteria than quantitative research for assessing its quality of measurement: authenticity, trustworthiness, credibility, dependability, and confirmability.

7.5. References

Babbie, Earl. 2010. *The Practice of Social Research.* 13th ed. Boston: Cengage Learning.

Bem, Sandra L. 1974. "The Measurement of Psychological Androgyny." *Journal of Consulting and Clinical Psychology* 42(2):155–62.

Bogardus, Emory Stephen. 1933. "A Social Distance Scale." *Sociology and Social Research* 17:265–71.

Coppock, Alexander. 2017. "Did Shy Trump Supporters Bias the 2016 Polls? Evidence from a Nationally-Representative List Experiment." *Statistics, Politics and Policy* 8(1):29–40. doi: 10.1515/spp-2016-0005.

Desmond, Matthew. 2016. *Evicted: Poverty and Profit in the American City.* New York: Broadway Books.

Enns, Peter K., Julius Lagodny, and Jonathon P. Schuldt. 2017. "Understanding the 2016 US Presidential Polls: The Importance of Hidden Trump Supporters." *Statistics, Politics and Policy* 8(1):41–63. doi: 10.1515/spp-2017-0003.

Fischer, Claude S., Michael Hout, Martín Sánchez Jankowski, Samuel R. Lucas, Ann Swidler, and Kim Voss. 1996. *Inequality by Design: Cracking the Bell Curve Myth.* Princeton, NJ: Princeton University Press.

Fisher, Gordon. 2008. "Remembering Mollie Orshansky—The Developer of the Poverty Thresholds." *Social Security Bulletin* 68(3):79–83.

Haghighat, Rahman. 2007. "The Development of the Brief Social Desirability Scale (BSDS)." *Europe's Journal of Psychology* 3(4). doi: 10.5964/ejop.v3i4.417.

Herrnstein, Richard J., and Charles A. Murray. 1994. *The Bell Curve: Intelligence and Class Structure in American Life.* New York: Free Press.

Hopkins, Daniel J. 2009. "No More Wilder Effect, Never a Whitman Effect: When and Why Polls Mislead about Black and Female Candidates." *Journal of Politics* 71(3):769–81. doi: 10.1017/S0022381609090707.

Iceland, John. 2013. *Poverty in America: A Handbook.* Berkeley: University of California Press.

Kurlaender, Michal, and Kramer A. Dykeman. 2019. *Predicting College Success: How Do Different High School Assessments Measure Up?* Stanford, CA: PACE.

Laloggia, John. 2019. "6 Facts about U.S. Political Independents." Pew Research Center. Retrieved December 12, 2022 (https://www.pewresearch.org/fact-tank/2019/05/15/facts-about-us-political-independents).

Marozzi, Marco. 2016. "Construction, Robustness Assessment and Application of an Index of Perceived Level of Socio-Economic Threat from Immigrants: A Study of 47 European Countries and Regions." *Social Indicators Research* 128(1):413–37. doi: 10.1007/s11205-015-1037-z.

Martin, Michel. 2008. "'Wilder (Bradley) Effect': Polling Myths or Truths." NPR, October 22.

Orshansky, Mollie. 1965. "Counting the Poor: Another Look at the Poverty Profile." *Social Security Bulletin* 28(1):3–29.

Payne, J. Gregory. 2010. "The Bradley Effect: Mediated Reality of Race and Politics in the 2008 U.S. Presidential Election." *American Behavioral Scientist* 54(4):417–35. doi: 10.1177/0002764210381713.

Rodwell, Mary K. 1998. *Social Work Constructivist Research.* New York: Garland.

Rothstein, Jesse M. 2004. “College Performance Predictions and the SAT.” *Journal of Econometrics* 121(1):297–317. doi: 10.1016/j.jeconom.2003.10.003.

Stiglitz, Joseph E. 2020. “GDP Is the Wrong Tool for Measuring What Matters.” *Scientific American*, August 1.

Swim, Janet K., Kathryn J. Aikin, Wayne S. Hall, and Barbara A. Hunter. 1995. “Sexism and Racism: Old-Fashioned and Modern Prejudices.” *Journal of Personality and Social Psychology* 68(2):199–214. doi: 10.1037/0022-3514.68.2.199.

Walch, Susan E., Sakkaphat T. Ngamake, Jacquelyn Francisco, Rashunda L. Stitt, and Kimberly A. Shingler. 2012. “The Attitudes toward Transgendered Individuals Scale: Psychometric Properties.” *Archives of Sexual Behavior* 41(5):1283–91. doi: 10.1007/s10508-012-9995-6.

Walker, Susan Noble, Karen Richert Sechrist, and Nola J. Pender. 1987. “The Health-Promoting Lifestyle Profile: Development and Psychometric Characteristics.” *Nursing Research* 36(2):76–82.

Watanabe, Teresa. 2019. “Grades vs. SAT Scores: Which Is a Better Predictor of College Success?” *Los Angeles Times*, December 22.

8. ETHICS



In 1971, police in Palo Alto, California, arrested nine young men for armed robbery and burglary and brought them to the city's police headquarters for processing. There, the men were warned of their Miranda rights, fingerprinted, and photographed while a local TV news crew filmed the arrest. Police cars with wailing sirens then transported the men to Stanford University. In the basement of Jordan Hall, three 7- × 10-foot cells had been constructed, each unlit and furnished only with three cots. Guards wearing uniforms and mirrored sunglasses met the prisoners, strip-searched them, and assigned them identification numbers. The prisoners were given smocks and stocking caps to wear over their otherwise naked bodies, and a chain was padlocked to their ankles. Guards addressed them only by their numbers (Social Psychology Network 2015).

The Stanford Prison Experiment had begun. The charges were fake, the arrests were staged, and the "prison" was a simulated detention center set up by Stanford psychology professor Philip Zimbardo in a basement section of the university's psychology building. The nine "prisoners" and nine "guards" were all male college students who had been recruited through local newspaper ads and were being paid $15 a day to be part of the two-week study. These participants—vetted beforehand to exclude those with mental or physical health problems or criminal backgrounds—had been randomly assigned to their "prisoner" or "guard" roles. The experiment's goal was to see if being thrust into positions of anonymous and impassive authority over other people would, by itself, make individuals abusive toward those under their control. As Zimbardo would say two decades later (O'Toole 1997):



College students playing the role of guards in the 1971 Stanford Prison Experiment wore mirrored sunglasses so that they could exercise their authority anonymously over the prisoners. *Stanford Prison Experiment, via PrisonExp.org*

> I had been conducting research for some years on deindividuation, vandalism and dehumanization that illustrated the ease with which ordinary people could be led to engage in anti-social acts by putting them in situations where they felt anonymous, or they could perceive of others in ways that made them less than human, as enemies or objects.

As Zimbardo expected, the experiment yielded "anti-social acts." In fact, the behaviors he observed were startlingly sociopathic. Although the guards were ordered not to cause physical harm, they quickly began humiliating their charges. When faced with widespread resistance early on, the guards sprayed prisoners with fire extinguishers. They seized the inmates' mattresses and stripped off their clothing. The discipline escalated to forcing uncooperative prisoners to do push-ups, spend time in solitary confinement, wear bags over their heads, and relieve themselves only in buckets in their cells.

By the fourth day, two of the prisoners had left the experiment—one after breaking down crying. On the fifth day, Zimbardo's colleagues and family and friends of the inmates came to visit the prison—the latter as part of a planned 10-minute session meant to simulate prison visitations—and some expressed horror at the inmates' treatment. On the sixth day, Zimbardo ended the study after the urging of psychologist Christina Maslach (his future spouse), who bluntly told him his research was immoral.



In the Stanford Prison Experiment, the nine prisoners lived in three 7- × 10-foot cells furnished only with cots. *Stanford Prison Experiment*, *via* PrisonExp.org

The Stanford Prison Experiment was a highly influential study published in a top psychology journal. It raised important questions about the extent to which ethical behavior is situational—that is, dependent on the specific social context that individuals confront, rather than their core principles. Zimbardo went on to write a bestselling book based on his study, *The Lucifer Effect: Understanding How Good People Turn Evil* (2007), and his research findings later became part of the public debate over the abuse of prisoners at the Abu Ghraib prison in Iraq, with Zimbardo testifying as an expert witness on behalf of one of the accused guards. Today, however, his study is most remembered not for its empirical contributions but for its ethical lapses.



On the fourth day, Prisoner 819 began crying in his cell. He refused an offer to speak to a priest and asked for a medical doctor instead. Eventually, the study's principal investigator, Philip Zimbardo, allowed the prisoner to leave the experiment. *Stanford Prison Experiment, via PrisonExp.org*

Ethics is about upholding standards of conduct. These standards can be personal—a matter of the religious beliefs or other moral principles you happen to adhere to—but they can also be specific to a particular field, such as sociology or the social sciences more broadly. In **Chapter 12: Experiments**, we'll talk about how the Stanford Prison Experiment was not actually an "experiment" and how it fell short of scientific rigor in a number of ways. In this chapter, however, we'll discuss it and other studies that raise questions about research ethics. Even studies with adequate or superior methodological approaches can still fail spectacularly when it comes to the respect and caring they show to their human subjects. As it turns out, the Stanford Prison Experiment happened to violate numerous tenets of research ethics—transgressions that went well beyond the study's treatment of the young men who were misled and abused as part of the experiment.

Morally dubious scientific studies can bring about tangible harms to research participants. Sometimes, individuals have died or been physically or emotionally scarred by participation in such studies. In social science research, the potential damage is more psychological, but researchers can also treat participants unethically by deceiving them or violating their privacy. And scientists can engage in immoral actions even when no human beings are directly hurt in the course of their work. For instance, throughout history, science has often been manipulated in unethical ways by individuals and organizations to promote and defend their agendas. Bias, misrepresentation, and falsification of data are serious transgressions that continue to harm public trust in the social sciences.

How exactly did the Stanford Prison Experiment fall short of ethical standards in research? Obviously, the psychological abuse that Zimbardo's "prisoner" subjects endured violated one such basic principle (**no harm to participants**). But the study also engaged in less sensationalistic research sins. For instance, the inmates were not told beforehand that they would be mock-arrested by Palo Alto Police (**informed consent**), and Zimbardo made it difficult for individuals to drop out of the experiment (**Video 8.1**). Reviewing the study, other scholars alleged that Zimbardo coached his "guards" to act more brutally than they otherwise would have, raising questions about bias in his study's results (Le Texier 2019).

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://viva.pressbooks.pub/sociology-research-methods/?p=256#oembed-1

Video 8.1. An Inauspicious Start to an Experiment. A "prisoner" in the Stanford Prison Experiment is mock-

arrested by a Palo Alto police officer—a part of the experiment that was not disclosed to participants beforehand as part of the informed consent process.

Over time, controversies over studies like the Stanford Prison Experiment have prompted reforms, with scientific disciplines establishing stronger professional norms and institutions forming oversight committees to regulate research. National and international organizations that represent scientists across scientific fields often maintain their own **professional codes of conduct**, setting standards for ethical work in each discipline. Universities and other organizations maintain internal bodies to enforce each institution's standards for research. This type of regulatory body, called an **institutional review board** (in the United States) or an *ethics committee* (in the United Kingdom), typically must approve any study by researchers who work with human subjects.

That said, it is important to note that ethical considerations in research are not always black and white. In the course of their work, researchers are inevitably faced with ethical choices. They must choose not to pursue certain approaches to their observation and analysis—even approaches that would make it easier to answer their research questions—whenever such choices are morally unacceptable. In many cases, these choices are obvious, but sometimes, they are not. Indeed, you may encounter scholars who defend some of the controversial studies mentioned in this chapter, perhaps with the argument that the ends justify the means. (As we will see, the U.S. government even prevented Axis scientists who conducted horrific human experiments during World War II from being prosecuted after the war, in part because of a view that some of the knowledge obtained so unethically was nonetheless useful.) For the sorts of studies that sociologists typically conduct, the consequences are usually not life or death, but there are, nonetheless, very real risks. Even well-intentioned social scientists have at times deceived or exploited the individuals and communities they were studying. Learning how scientists in the past have strayed from ethical principles is essential so that you—as a good sociologist and a good human being—avoid a similar complacency or callousness.

Opening chapter image credit: *Katrin Bolovtsova, via Pexels. Adapted by Bizhan Khodabandeh.*

8.1. Subject to Harm: Coercion, Deception, and Other Risks for Participants



A human subject in the Tuskegee Syphilis Study. *U.S. Centers for Disease Control and Prevention, via Wikimedia Commons*

Learning Objectives

1. Define human subjects research and describe its differences with non-human subjects research.
2. Describe and provide examples of the various ways that research can harm human subjects.

Ideally, all types of research are limited by various sets of ethical concerns. However, research on human subjects is regulated much more heavily than research on nonhuman subjects. According to the U.S. Department of Health and Human Services (2017:7260), a **human subject** is a "living individual about whom an investigator . . . conducting research":

1. "Obtains information or biospecimens through intervention or interaction with the individual, and uses, studies, or analyzes the information or biospecimens"; or
2. "Obtains, uses, studies, analyzes, or generates identifiable private information or identifiable biospecimens."

A couple of things are worth noting about the federal policy that oversees human subjects. First, it explicitly states that its regulations apply to anyone conducting research—"whether professional or student." Second, even if the researcher does not interact directly with the individual, the recording of any identifying details—which may undermine the individual's privacy and safety—necessitates greater care on the part of the researcher to uphold ethical practices.

Early trials of medical vaccinations provoked wide-ranging public debates over the ethics of conducting potentially dangerous research on human subjects (Rothman 1987). In 1796, scientist Edward Jenner exposed an eight-year-old boy to a virus related to smallpox in order to develop a vaccine for the devastating disease (Riedel 2005). Jenner believed the dangers posed to the child were outweighed by the need to find a cure—an endeavor that he argued would benefit the whole of humanity. After the exposure, the boy fell sick with a fever and then chills, but within days, he felt much better. Then Jenner exposed the child again—this time, to smallpox. This time, nothing happened: the vaccination worked. While Jenner was exceedingly lucky that his human subject in this case did not suffer long-term harm, things could easily have turned out far worse.

From a modern perspective, subjecting another human being to this level of danger would be unacceptable and unnecessary for a vaccine trial. It would violate a core principle of research ethics: to do *no harm to participants*. In rare cases, putting participants at risk of harm can be justified—such as when a new and experimental medical treatment is being tested—but the benefits of such risks must be substantial, and the approach taken should be a last resort, used only when alternatives have failed or are not possible. For this reason, new vaccines and other treatments are typically tested on animal subjects (which themselves deserve ethical consideration) before researchers move onto humans. Even then, there are hard lines that prevent researchers from endangering participants.

Starting largely in the twentieth century, public anger and horror over serious breaches in research ethics led to the development of rules—established both by government laws and professional norms—to rein in bad scientific behavior. Yet even when such rules existed, there was little accountability. At the end of World War II, Nazi doctors and scientists were put on trial by the Allies for conducting human experimentation on prisoners. Germany had already established regulations specifying that human subjects must clearly and willingly consent to their participation in medical research, a principle known as *informed consent*. Nazi scientists utterly ignored these regulations, however. They saw the human subjects forced to participate in their gruesome experiments—concentration camp inmates—as less than human and thus not deserving of protection.[1]

With no checks on their capacities for cruelty, Nazi scientists tortured and murdered thousands—not just political prisoners, but also Jewish, Polish, Roma, and LGBTQ individuals singled out as especially undesirable (Weindling et al. 2016). On an even larger scale, Japan's infamous Unit 731 engaged in lethal experiments—ranging from exposing patients

1. In addition to Jewish communities, Nazi ideology dehumanized those with mental and physical disabilities, ethnic minorities, and those who did not conform to heteronormative standards. Richard J. Evans (2006) discusses how the Nazis promoted forced sterilization and involuntary euthanasia, while George J. Annas and Michael A. Grodin (1995) document how the unethical practices of Nazi doctors and scientists informed the development of the Nuremberg Code.

to frostbite, diseases, and biological weapons, to performing vivisections and amputations—which were conducted on men, women, and children, most of them Chinese and Russian.



A 1944 photo of a Romani prisoner at the Dachau concentration camp who was forced to serve as an unwilling human subject. Nazi scientists there were testing methods of converting seawater—lethal because of its salt content—into drinkable water. *U.S. National Archives and Records Administration, via the U.S. Holocaust Memorial Museum*

After Germany surrendered, the Allies prosecuted some of the doctors who had conducted human experiments at Auschwitz, Dachau, and other concentration camps. One result of these tribunals was the creation of the **Nuremberg Code**, 10 research principles designed to guide doctors and scientists who conduct research on human subjects (see the sidebar **The Nuremberg Code**). This code would inspire later rules adopted by numerous national governments and professional bodies. As for Japan's even more egregious research program, the U.S. government deliberately concealed its knowledge of Unit 731's activities so that war crimes prosecutors were unable to try the scientists responsible. U.S. general Douglas MacArthur secretly gave immunity to the researchers to gain full access to what was seen as valuable data—in addition to keeping Japan's horrifically acquired knowledge about biological weapons away from the Soviets.

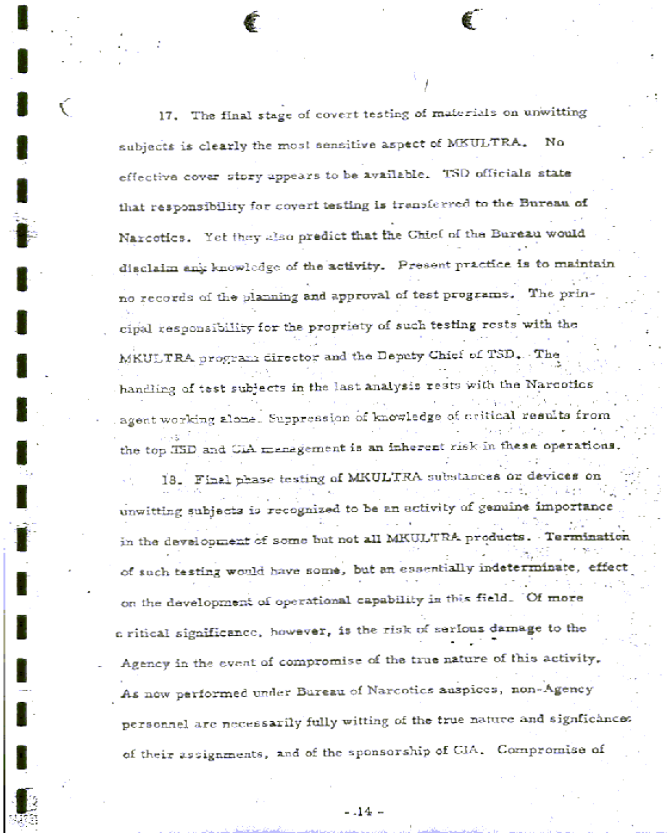
Even as Allied nations learned about the extent of the Axis's human experimentation and established ethical guidelines in response, they remained hypocritically willing to ignore the transgressions of their own researchers—including those working for government agencies. Between 1932 and 1972, for example, the U.S. Public Health Service oversaw the Tuskegee Syphilis Study, one of the most notorious medical studies ever conducted. The researchers recruited 600 African American sharecroppers from impoverished areas of Macon County, Alabama, promising them free medical care to be part of the study. Among the recruited participants, 399 men had latent syphilis but were not told of their condition. The researchers wanted to study the effects of the disease when it was left untreated, so they set up an elaborate ruse, giving the men false treatments for "bad blood" while comparing their outcomes to those of the study's control group, the 201 men without syphilis. The researchers continued to deny the men effective care even after penicillin became a widely available treatment in the 1940s.

A leak to the press finally forced officials to shut down the experiment in 1972. By then, the failure to provide actual treatment over four decades had already brought about the deaths of 128 of the men from syphilis or related complications.



U.S. Army scientist Frank Olson (played by Peter Sarsgaard) receiving a psychological test as part of the Central Intelligence Agency's mind control program MKUltra in the film *Wormwood* (2017). *Netflix*

Inspired by experiments on mind control conducted by Japanese and Nazi doctors during World War II, the U.S. Central Intelligence Agency (CIA) embarked in 1953 on a two-decade research initiative called Project MKUltra (also known as MK-ULTRA). According to journalist Stephen Kinzer (2019), the covert CIA project was not only "roughly based" on those earlier experiments, "but the CIA actually hired the vivisectionists and the torturers who had worked in Japan and in Nazi concentration camps to come and explain what they had found out so that we could build on their research" (Gross 2019). For example, Nazi scientists who had experimented with mescaline at the Dachau concentration camp were brought aboard as advisors, Kinzer reports. The CIA also recruited Nazi doctors to lecture officers about "how long it took for people to die from sarin"—a poison gas the Nazis had also studied.



17. The final stage of covert testing of materials on unwitting subjects is clearly the most sensitive aspect of MKULTRA. No effective cover story appears to be available. TSD officials state that responsibility for covert testing is transferred to the Bureau of Narcotics. Yet they also predict that the Chief of the Bureau would disclaim any knowledge of the activity. Present practice is to maintain no records of the planning and approval of test programs. The principal responsibility for the propriety of such testing rests with the MKULTRA program director and the Deputy Chief of TSD. The handling of test subjects in the last analysis rests with the Narcotics agent working alone. Suppression of knowledge of critical results from the top TSD and CIA management is an inherent risk in these operations.

18. Final phase testing of MKULTRA substances or devices on unwitting subjects is recognized to be an activity of genuine importance in the development of some but not all MKULTRA products. Termination of such testing would have some, but an essentially indeterminate, effect on the development of operational capability in this field. Of more critical significance, however, is the risk of serious damage to the Agency in the event of compromise of the true nature of this activity. As now performed under Bureau of Narcotics auspices, non-Agency personnel are necessarily fully witting of the true nature and significance of their assignments, and of the sponsorship of CIA. Compromise of

- 14 -

A declassified document from the CIA's Project MKUltra, which studied mind control techniques. The CIA destroyed most of the files related to this clandestine research program before investigators could learn the full extent of the abuses perpetrated on—as this document describes—often "unwitting subjects." *Wikimedia Commons*

As part of MKUltra, CIA officers set up secret detention centers throughout Europe and East Asia where they experimented on captured enemy agents, testing not only drugs but also electroshock, extremes of temperature, sensory isolation, and harsh interrogation techniques meant to break down the ego. Government investigations into MKUltra finally put an end to the CIA program in the 1970s, although the CIA successfully destroyed most of the documents related to the project before investigators could obtain them.

Both the Tuskegee study and the CIA's Project MKUltra caused unconscionable harm to participants and failed to obtain their informed consent. The consequences of these betrayals of basic ethical principles continue to reverberate today. The ensuing scandal over the Tuskegee experiment led to the creation of the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research (discussed further in the next section), a U.S. federal agency devoted to regulating research. In spite of the reforms that ensued, the Tuskegee experiment is still brought up—many decades later—as a critical historical context for why many African Americans have a deep distrust of the medical establishment (Katz et al. 2008). MKUltra prompted a series of lawsuits against the federal government, with research participants and their survivors calling out the CIA's flagrant use of deception and coercion. And thanks to the agency's brazen document destruction, Project MkUltra still feeds wide-ranging conspiracy theories about the federal government to this day.

The Nuremberg Code



Defendants in the dock during the war crime trials held by the International Military Tribunal in Nuremberg, Germany (November 1945). *Raymond D'Addario, via Wikimedia Commons*

In 1947, the Nuremberg war crimes tribunal laid down 10 principles for "permissible medical experiments," which would become known as the Nuremberg Code (Vollmann and Winau 1996:1448):

> The great weight of the evidence before us is to the effect that certain types of medical experiments on human beings, when kept within reasonably well-defined bounds, conform to the ethics of the medical profession generally. The protagonists of the practice of human experimentation justify their views on the basis that such experiments yield results for the good of society that are unprocurable by other methods or means of study. All agree, however, that certain basic principles must be observed in order to satisfy moral, ethical and legal concepts:

1. The voluntary consent of the human subject is absolutely essential: This means that the person involved should have legal capacity to give consent; should be so situated as to be able to exercise free power of choice, without the intervention of any element of force, fraud, deceit, duress, overreaching, or other ulterior form of constraint or coercion; and should have sufficient knowledge and comprehension of the elements of the subject matter involved as to enable him to make an understanding and enlightened decision.

2. The experiment should be such as to yield fruitful results for the good of society, unprocurable by other methods or means of study, and not random and unnecessary in nature.

3. The experiment should be so designed and based on the results of animal experimentation and a knowledge of the natural history of the disease or other problem under study that the anticipated results will justify the performance of the experiment.

4. The experiment should be so conducted as to avoid all unnecessary physical and mental suffering and injury.

5. No experiment should be conducted where there is an a priori reason to believe that death or disabling injury will occur; except, perhaps, in those experiments where the experimental physicians also serve as subjects.

6. The degree of risk to be taken should never exceed that determined by the humanitarian importance of the problem to be solved by the experiment.

7. Proper preparations should be made and adequate facilities provided to protect the experimental subject against even remote possibilities of injury, disability, or death.

8. The experiment should be conducted only by scientifically qualified persons. The highest degree of skill and care should be required through all stages of the experiment of those who conduct or engage in the experiment.

9. During the course of the experiment the human subject should be at liberty to bring the experiment to an end if he has reached the physical or mental state where continuation of the experiment seems to him to be impossible.

10. During the course of the experiment the scientist in charge must be prepared to terminate the experiment at any stage, if he has probable cause to believe, in the exercise of the good faith, superior skill and careful judgment required of him that a continuation of the experiment is likely to result in injury, disability, or death to the experimental subject.

Ethical Questions about Social Science Research



An iconic sketch of Ted Kaczynski, the Unabomber, which the Federal Bureau of Investigation (FBI) widely distributed in the 1990s while Kaczynski was one of the agency's most wanted fugitives. During college, Kaczynski participated in a long-running Harvard psychological study in which he and other students were regularly verbally abused to examine their reactions to extreme stress. *Jeanne Boylan (FBI)*

When regulators talk about putting safeguards on scientific research, they often distinguish between *biomedical* and *behavioral* research. The latter group includes the social sciences in general and sociology in particular. Unethical biomedical studies are especially troubling because of their potentially lethal consequences. Yet morally dubious research in the social sciences also occurs. Even if human subjects face no mortal threat, they can be subjected to distress or other psychological harm as a result of their participation. For this reason, governmental agencies, institutional review boards (IRBs), and other regulatory bodies routinely scrutinize behavioral studies for possible risks to their human subjects, even though the stakes of such research are generally lower than those for biomedical studies.

Among the various social scientific methods of research, experiments often entail the most intrusion into the lives of human participants, given that by definition they involve not just observing social interactions, but also manipulating those interactions to study the consequences (we'll have more to say about the differences between experimental and observational data in **Chapter 12: Experiments**). As a result, some of the most well-known instances of problematic research in the social sciences comes from the field of psychology, which relies heavily on lab experiments.

Starting in 1959, Harvard psychologist Henry Murray began a multiyear experiment on acute stress that involved 22 undergraduates at his university. The students were told they would debate philosophy with a fellow student. In preparation, they wrote essays about their closely held beliefs and aspirations. Without their knowledge, Murray had a litigator look over their essays and then sit down with each student to grill and belittle them. Electrodes attached to the subjects measured their physiological responses. The critic's "vehement, sweeping, and personally abusive" attacks were intended to break down their egos and generate extreme distress, Murray said. The students then repeatedly watched film recordings of their frightened and angry reactions during the abuse sessions.

One of the subjects in Murray's study was a 17-year-old sophomore named Ted Kaczynski, a brilliant but socially awkward math student who had skipped 11th grade to attend Harvard a year early. Journalist Alston Chase (2000) describes the experiment as "purposely brutalizing" and argues that it set Kaczynski down a dark path. Until Murray's study ended in 1962, Kaczynski attended weekly sessions—200 hours in total—so that Murray could measure the impact of sustained abuse and humiliation on his subjects.

Kaczynski graduated from Harvard, obtained his PhD, and began a promising career as a math professor at the University of California, Berkeley. But in 1969, he abruptly resigned, having become—in his department chair's words—"almost pathologically shy." He moved out to a remote cabin in Montana to live as a survivalist. There, Kaczynski came to see technology and industrialization as the enemies of human freedom and dedicated himself to avenging the harm they had wrought on the natural world. He began sending mail bombs across the country, at first targeting engineering professors and the president of United Airlines—which prompted the FBI to use UNABOM (Universities and Airlines) as his case file name. The Unabomber was finally captured in 1996, but not before his terror spree killed three

people, injured and maimed 23, and forced the *Washington Post* to give into his demand to publish his 35,000-word anti-technology manifesto.

Perhaps the most cited (and even parodied) example of unethical psychological experimentation is a series of studies conducted by Stanley Milgram (1963, 1974) at Yale University to understand obedience to authority. Somewhat ironically, given how much criticism the ethics of his study would face, Milgram's goal was to explore whether good people could be convinced to do evil things—a moral question made especially salient by the high-profile trial underway around that time of Nazi war criminal Adolf Eichmann. As Milgram would later write in a 1974 book on his experiments, "Could it be that Eichmann and his million accomplices in the Holocaust were just following orders? Could we call them all accomplices?"



A lurid example of unethical research, the Milgram experiment—which featured "teachers" pressing buttons to shock "learners"—has entered pop culture. *Walt Disney Company, via IMDb*

In 1961, Milgram recruited 40 young and middle-aged men with diverse occupations and levels of education as his research subjects. They took on the role of "teachers" in an experiment that was supposed to test the ability of a "learner" to memorize word pairs. The teacher and learner were separated by a wall so that they could not see each other but could hear what the other said. In each iteration of the experiment, the learner would be strapped to an electric chair with a shock generator. When the learner failed to remember a word pair, the subject was told to press a button to shock them. The severity of the shocks began at a low voltage and then increased with each new error. The machine could deliver shocks between 15 and 450 volts—explicitly labeled from "Slight Shock" to "Danger: Severe Shock" on the generator. If the subject hesitated to deliver a shock, the attending researcher—dressed in a white lab coat to signal their authority—would insist that they continue, becoming more demanding if the subject did not obey. As the shocks increased in intensity, learners would cry out, at some points banging against the wall in protest (if asked, the researcher would reassure the teacher that the shocks were not causing permanent damage). In later versions of the experiment, learners would even plead for mercy and reveal they had a heart condition. At the highest voltages, the learners would fall silent.

In reality, the electric shocks were not real, and the "learners" were actually actors (in lab lingo, **confederates** of the researchers) who were only pretending to be in pain. Milgram had expected only a small minority of participants to apply a severe shock, in line with the predictions of the students and other faculty he polled before the experiments. He was startled to find that every participant in his study went up to 300 volts, and two-thirds allowed the shocks to reach the maximum 450 volts. All of the participants paused the experiment to raise questions about it, but none of them—even those who refused to proceed to the highest voltage—demanded that the study be ended.

Although the learners only pretended to be in pain, many of Milgram's research participants experienced emotional distress during the experiment, with some falling into fits of nervous laughter and even seizures (Ogden 2008). Milgram later said that most of the participants expressed gratitude for being part of the study. Nevertheless, some of his peers condemned the study for the psychological harm it potentially caused both at the time of the experiment and over the long term. The realization that you are willing to administer painful and even lethal shocks to another human being just because someone who looks authoritative has told you to do so might indeed be traumatizing—even if you later learn that the shocks were faked. After Milgram's original 1963 paper came out, a number of replications and variations of his study were conducted (including by Milgram himself) and arrived at similar results. Yet the fierce criticism of the

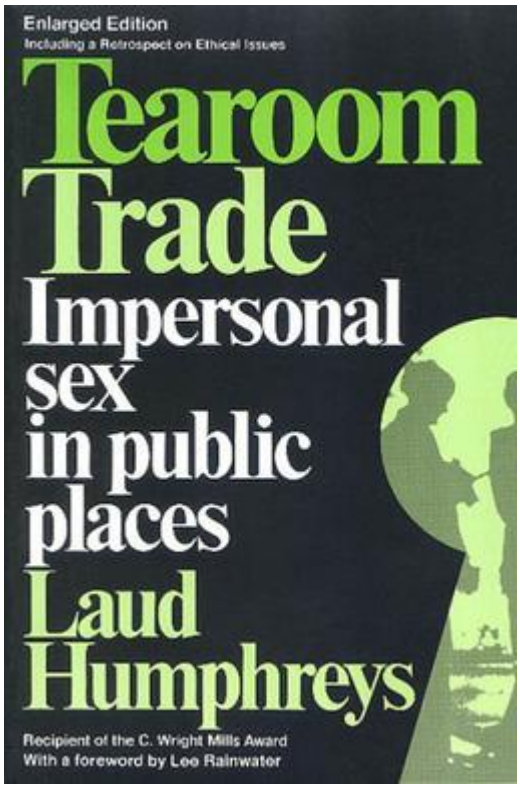
morality of Milgram's experiments led to widespread discussion among psychologists about the potential risks of their work and ultimately brought about an overhaul of the discipline's prevailing ethical standards.

While perhaps less widely known, a number of nonexperimental studies have also led to ethical scandals. Around the same time as Milgram's experiments, sociology doctoral student Laud Humphreys (1970) was collecting data for his dissertation research, which employed the qualitative method known as **ethnography** (discussed in **Chapter 9: Ethnography**). Humphreys was studying the "tearoom trade," the practice of men engaging in anonymous sexual liaisons in public restrooms. He was trying to understand who the men were and why they sought out these encounters. During his research, he conducted **participant observation** by serving as a "watch queen," the person who keeps an eye out for police in exchange for being able to watch the sexual encounters.

Humphreys spent months with his subjects, getting to know them and learning more about the tearoom trade. Without their knowledge, he wrote down their license plate numbers as they pulled into or out of the parking lot near the restroom he was observing. With the help of some collaborators who had access to motor vehicle registration information, Humphreys used the plate numbers to find out the names and home addresses of his research subjects. Again without their consent, he went to their homes unannounced, disguising his identity and claiming he was a public health researcher. Under that pretense, he interviewed the men about their lives and their health, never mentioning his previous observations.

As you probably surmised, Humphreys's study violated numerous principles of research ethics. He failed to identify himself as a researcher who was studying sexual encounters. He intruded on his participants' privacy by obtaining their names and addresses and approaching them in their homes. He deceived them by claiming he was doing a public health study and not acknowledging any connection to his earlier observations. In spite of all these ethical breaches, the research that Humphreys conducted arguably did a lot of good. For one thing, it challenged myths and stereotypes about who was engaging in these sexual practices. Humphreys learned that more than half of his subjects were married to women, and many did not identify as gay or bisexual. This knowledge helped decrease the intense social stigma at the time surrounding "men who have sex with men," an umbrella term coined to include the latter group (Lehmiller 2017). Humphreys's study even reportedly prompted some police departments to discontinue their raids on public restrooms and other places where anonymous sexual liaisons were taking place.

After its publication, however, Humphreys's work became controversial, opening up a still ongoing discussion in the discipline about the appropriate amount of informed consent and protection of privacy that ethnographers should ensure to conduct their work both effectively and ethically. Some scholars called the tearoom trade study unethical because Humphreys misled his participants and put them at risk of losing their families and their positions in society, given the stigma associated with gay sex (Warwick 1982). His defenders pointed out that the sexual encounters he observed occurred in public places with no legitimate expectation of privacy. And the benefits of Humphreys's research—namely, the way it dispelled myths about the tearoom trade specifically and human sexual behavior more generally—outweighed any harm it may have caused, they said (Lenza 2004).



Laud Humphreys's book *Tearoom Trade* (1970, 2008) sparked controversy because of the subterfuge that Humphreys employed in gathering personal details about men he observed having anonymous sex in public restrooms.

Looking back almost four decades after his study was published, Humphreys (2008) maintained that none of his research subjects had actually suffered because of his study, and that the "demystification of impersonal sex" had been of benefit to society. Over the years, however, Humphreys had also come to the conclusion that he should have conducted the second part of his study differently. Instead of writing down license plate numbers to track down his subjects at home and lying to them in the guise of a public health researcher, Humphreys could have spent more time in the field, he wrote. By doing so, he could have gained the trust of those he was studying and enlarged the pool of his informants. Then he could have divulged his true identity and sought their consent for his use of their data.

Today, sociologists would find it extremely difficult to conduct a project similar to Humphreys's. When he did his fieldwork, an effective regulatory framework for policing scientific research was not yet in place, and sociology and other disciplines had not yet defined their standards for ethical research in a clear and comprehensive fashion. As we will discuss later in the chapter, these regulations and professional norms mean that researchers must design their studies in specific ways to protect their human subjects. At the same time, we should note the gray areas that exist even with today's more enlightened approach. Ambitious and important studies can create risks for subjects, but they can also be defended with ends-justify-the-means arguments—namely, that the benefits of the research outweigh its risks, like Humphreys's defenders said his study did. Where to draw the line is a topic of constant debate, and the consensus may shift with each new generation of researchers—as our modern-day horror over studies that seemed appropriate to researchers decades ago may remind us.

Key Takeaways

1. Because in sociology we deal primarily with human research subjects, sociologists face particular challenges in conducting research ethically.

2. The history of science (including the social sciences) includes numerous examples of highly problematic research that harmed or otherwise negatively impacted human subjects.

Exercises

1. Do you believe Humphreys's research methods were ethically wrong? Or were they justifiable because of their larger (arguably positive) impacts?
2. How would you change Humphreys's research methods to make them more ethically acceptable?

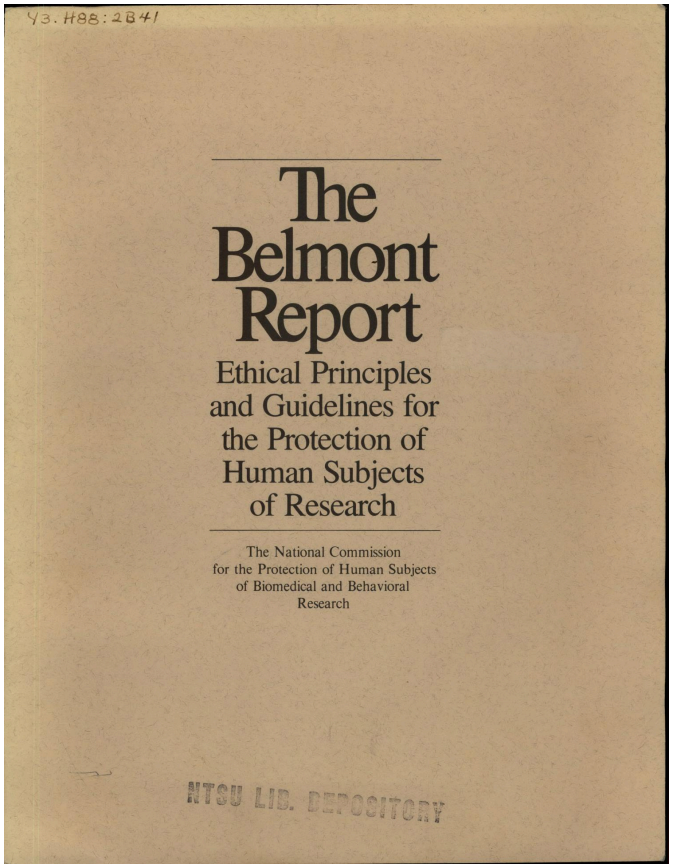
8.2. The Belmont Principles: Respect for Persons, Beneficence, and Justice

Learning Objectives

1. Identify the key principles of ethical research and how exactly they affect the work that scientists do.
2. Explain why sociologists might want to give their research participants protections such as anonymity and confidentiality.

In the United States, the foundational document guiding government regulations and professional norms regarding research ethics is the **Belmont Report**, which grew out of the federal government's efforts to establish a regulatory framework for research in the 1970s. As part of the 1974 National Research Act, Congress required all institutions receiving federal support to establish IRBs to oversee research projects and protect the rights of human subjects. The legislation also launched the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, which in 1979 produced the *Belmont Report*—named after the Maryland manor where the commission met. Inspired by the Nuremberg Code, the *Belmont Report* put forward three overarching ethical principles for research on human subjects: *respect for persons*, *beneficence*, and *justice* (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research 1979). Let's break down the implications of these three principles (you can also read the full text of the report online):

> **1. Respect for persons:** Individuals should be treated as autonomous agents able to make their own decisions. This principle implies informed consent—honoring your research participants' autonomy means that you should tell them about what you are studying, how you are studying it, and what risks and benefits their participation entails. Showing respect for participants also means you should safeguard their privacy, a topic we discuss further in regards to **protection of**



In 1979, the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research released the *Belmont Report*. Named after the conference center where the commission drafted the report, the report laid out three key principles for ethical research: respect for persons, beneficence, and justice. *University of North Texas Digital Library*

identities. Finally, the principle of respect for persons applies to groups with diminished autonomy—such as children and those with disabilities—for whom scientists must show special care when involving them in research.

2. Beneficence: Although the word "beneficence" is typically associated with unconditional acts of kindness and charity, in a research context, it is a moral imperative, the *Belmont Report* argues. Beneficence in this sense involves two principles: (a) "do not harm" (in line with the *no harm to participants* principle we have previously discussed) and (b) "maximize possible benefits and minimize possible harms." When researchers act with beneficence, they respect the decisions of their human subjects and secure the well-being of these individuals.

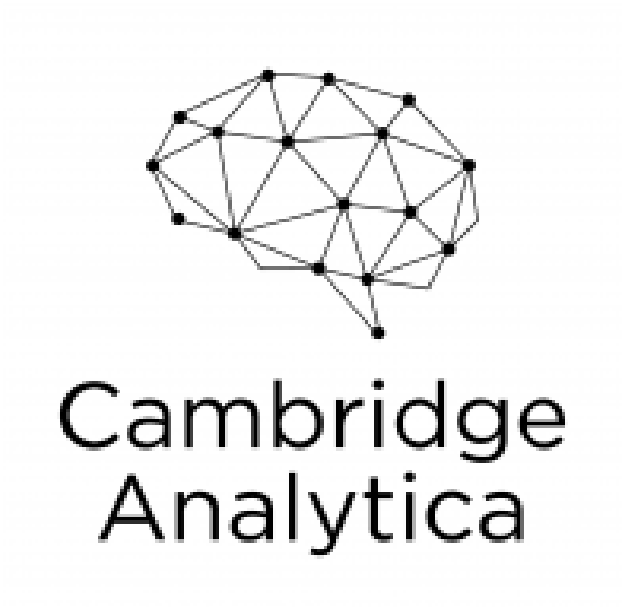
3. Justice: This principle focuses on who should receive the benefits of research and bear its burdens. An injustice occurs when some benefit to which a person is entitled is denied without good reason, or when some burden is imposed unduly or arbitrarily. The tenet of justice reinforces the need for informed consent: participants must receive clear information not only about the purposes of the study but also its risks and rewards.

We've already discussed the principle of no harm to participants that is explicit in the Belmont's beneficence tenet and was egregiously lacking in many of the biomedical and social scientific studies we've described thus far. In this section, we elaborate on the key concepts of *informed consent* and *protection of identities* that follow from the overarching principles in the *Belmont Report*.

Informed Consent

What does it mean to treat our human subjects as autonomous agents—that is, individuals who are capable of making decisions for themselves? It means that we seek their voluntary participation based on informed consent. In any legitimate scientific study—certainly any study an IRB is willing to approve—we must have formal procedures in place to ensure that all human subjects involved in the research freely choose to participate. Their consent should also be "informed," meaning that they should be given understandable and accurate information about the study, not some vague or misleading statements about it.

These formal disclosures are usually included in a **consent form** that researchers provide to subjects before the start of the study. In some cases, subjects are asked to sign the consent form indicating that they have read it and fully understand its contents. In other cases, subjects are simply provided a copy of the consent form, and researchers are responsible for making sure that subjects have read and comprehend its terms before starting any data collection. In either case, researchers should go over the form with their subjects at length—highlighting important sections of it, giving them ample time to read and review it, patiently addressing any questions or concerns they have about it, and providing them with a copy of the form to take home.



Facebook was embroiled in controversy after journalists uncovered evidence that the political consulting firm Cambridge Analytic had used the social media platform to gather personal details about Facebook users without their consent, which the company used to help the 2016 Trump presidential campaign target its political advertising. *Wikimedia Commons*

What sort of information should be laid out in a consent form and discussed during the consent process? First, researchers should describe what the general aims and goals of the study are, who is conducting and sponsoring the research, and how the research will be conducted—including what participants will do, and for how long. They should clearly communicate the potential risks associated with it, which obviously would include any psychological or safety risks but also any dangers to privacy or reputation. The informed consent process should also cover the possible benefits of participation—anything from monetary compensation to the opportunity for participants to share their personal views on an important issue. Finally, researchers must describe how they will protect the identities of their participants (covered in the next section) and whom they should contact for additional information about the study or their rights as study participants. This informed consent process should happen before data collection begins so that the knowledge can help participants decide whether or not they wish to be involved.

There are important practical implications of this general approach of informed consent:

- We cannot force anyone to participate in our research.
- We cannot include a person in our study without their knowledge or consent.
- We need to provide *clear* and *accessible* information to potential participants about the study they are being asked to be involved in.
- We cannot ask potential participants as a condition of their participation to not pursue legal recourse if the study somehow ends up harming them.

These last two points bear further discussion. In providing information about the study to potential participants, we need to describe the study in layperson's terms—not with dense scientific jargon—and we need to make their rights and any potential risks clear and concrete to them, rather than concealing them in ambiguous language or legalese. We also need to make sure that every potential participant fully understands what is being asked of them. This is an especially important consideration for more vulnerable populations like minors or immigrants, who may have particular trouble comprehending the parameters of the research study they are agreeing to.

On the final point, it is essential not to confuse informed consent with any strategy to avoid liability. Human subjects should not be forced or expected to release a researcher or other parties from legal consequences should something go wrong while the research is being conducted. And in giving their informed consent, subjects should not be asked to waive any of their legal rights.

Informed consent is not always necessary when conducting research. Remember how the U.S. federal government's definition of research on "human subjects" that we quoted at the beginning of this chapter stated that it involved "intervention or interaction with the individual" or the use or analysis of "private information." When we study public speeches given by other human beings, or published books written by them, or publicly released videos or audio recordings created by them, we are not technically conducting research on "human subjects"; instead, we are studying the publicly shared information that those individuals produced. In these cases, we do not need to ask those individuals for permission to study their public communications, which required no intervention or interaction with the individual for us to obtain. Likewise, when we observe public spaces—street corners, for instance—the normal assumption is that no consent is needed to write about or otherwise record what one observes (although there are caveats to this general view, as the uproar over Laud Humphreys's study of "public" restrooms suggests).

At the same time, it is not always clear what is "private" or "public" information. An everyday example of why researchers should insist on informed consent is the widespread and arguably intrusive analysis of user data for research conducted by companies and advocacy groups. Whenever you sign up for a new app or online service, you typically have to wade through thousands of words of legal disclosures—the dreaded end-user license agreement (EULA)—and click on the "Agree" button before you can upload your cat pictures. Embedded in these disclosures is often an acknowledgment that the provider of the service is using your data for research purposes. They want to know the demographics and habits of their users so that they can improve their service or hawk advertising to marketers who want to reach specific audiences. They often sell their data—including names, mailing addresses, email addresses, and other identifying information—to third parties who could use it for anything from sending you unsolicited emails to developing strategies to encourage (or discourage) you from voting.

In 2018, Facebook faced a major scandal when news reports revealed that a British consultancy called Cambridge Analytica had harvested the data of 87 million Facebook users and—without their knowledge or consent—had used their data to develop "psychographic" profiles that helped political organizations target more effective advertisements at supporters and swing voters across media platforms. The company's clients included the 2016 presidential campaign of Donald Trump, a Super PAC supporting the Trump campaign, and the 2016 British campaign in favor of a referendum vote to leave the European Union. Individuals close to Trump helped run and fund the company. The data harvesting first occurred through a popular app, "This Is Your Digital Life," that succeeded in getting hundreds of thousands of users to complete a paid survey said to be part of an academic study. These users agreed to language asking for their consent before taking the survey, yet Facebook allowed Cambridge Analytica to use their personal data for purposes other than what they actually consented to. The company was also able to pull data about the users' friends—thereby expanding the company's sample to encompass millions of people on Facebook. As a result of the scandal, Cambridge Analytica folded, and Facebook wound up paying billions of dollars in fines levied by U.S. and UK regulators.

While the deception uncovered in the Facebook/Cambridge Analytica data scandal was shamelessly brazen, it is clear that the "consent" process that companies typically use—an online agreement spelling out the data collection and its uses in fine print—is still a pale imitation of what an informed consent procedure looks like (or at least should look like) in

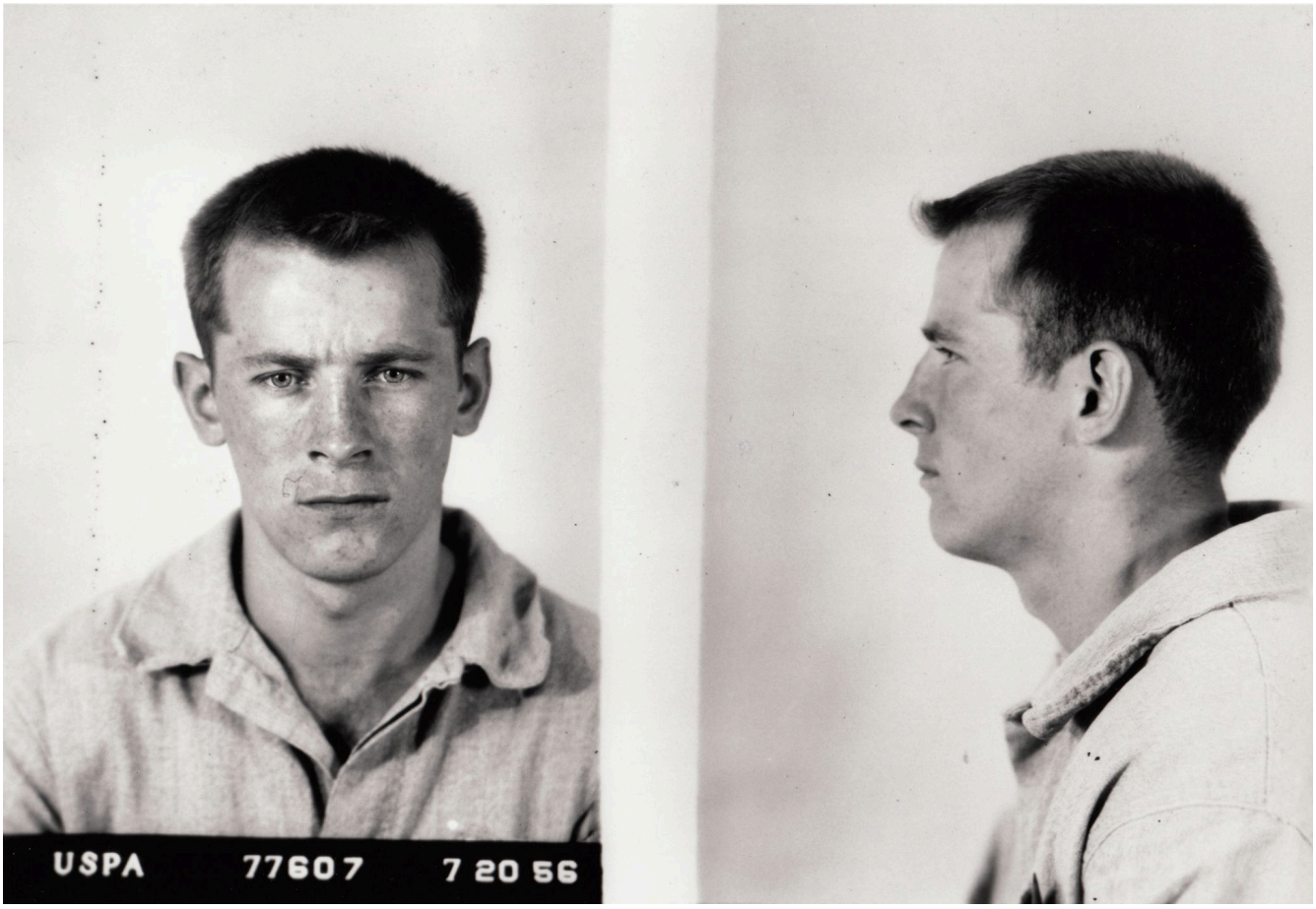
academic research. An ethical researcher needs to be utterly clear about the risks and benefits of the research, convey that information in an accessible and straightforward way, and fully address any questions or concerns without coercion of any kind—which is often the opposite of what the EULA's impenetrable legalese and "click to proceed" pressure tactics entail.

That said, there are exceptional cases when ethical researchers can justify not fully informing their subjects of the purpose or specific procedures of a study at its outset. In these situations, researchers may be allowed to speak vaguely about their research or even mislead their participants outright. IRBs may approve such strategies when there are serious concerns that participants who learn too much about the study beforehand will subsequently alter their responses or behavior, hampering efforts to get an accurate picture of the phenomenon being examined. For instance, suppose the purpose of a study is to examine to what extent subjects abandon their own views to conform with "groupthink." To measure this phenomenon in the lab, researchers convene a group of participants who are encouraged to listen to each other's opinions on a topic before voicing their own. In this research context, disclosing the study's purpose beforehand might sensitize subjects to the idea of groupthink and—consciously or not—make them act differently than they would have without that information.

Not being fully upfront about the nature of a study is often unavoidable if the researcher wants to conduct a valid analysis of a highly stigmatized phenomenon. For that reason, ethnographers sometimes conduct covert observations—much like undercover officers do when conducting drug busts—so that the people they observe act naturally (and perhaps tolerate the researcher's presence to begin with, as we discuss further in the next chapter). The social stigma associated with certain behaviors or beliefs makes informed consent particularly tricky for in-depth interviews. In this situation, the researcher usually knows the identities of the participants and must rely on their self-reporting about how they think or feel. Given that respondents can choose how they express themselves in these situations, they may tailor their answers to show themselves in a more favorable light. Specifically, they may hide less desirable behaviors or attitudes while exaggerating more desirable ones. In these cases, so-called **social desirability bias** will distort the researcher's understanding of what's actually going on. For instance, if a sociologist is conducting in-depth interviews with subjects to suss out their underlying racial biases, telling them the actual point of the study will prompt a good number of respondents to provide answers that paint themselves more positively than they actually are. Very few people will comfortably acknowledge that they are unrepentant bigots; at the very least, they will use euphemisms or code words and otherwise dial back the attitudes they might state bluntly to friends and family.

Weighing the pros and cons of *not* providing informed consent is a tricky matter, and sociologists and IRBs frequently have divergent opinions about what sorts of details can be justifiably concealed. Even if researchers have a strong case for keeping certain information from participants, they can and should conduct a debriefing session immediately following the data collection process. This ensures that the human subjects are eventually informed, even if it cannot occur before they participate in the study. In a **debriefing**, the researchers explain the true purpose of the study and why any deception was necessary. They also provide a full accounting of the potential risks or harm that the participant might have experienced during the experiment or might still suffer in the future. In all cases, the researchers need to provide ways to contact them or others who can offer assistance while the participants process their experiences afterward.

When obtaining informed consent from our potential participants, we should also keep in mind that some people in our pool of recruits may not be as able as others to give their genuine consent to participate in research. Here, we should revisit the *Belmont Report*'s principle of "respect for persons." Recall that this principle not only aims to treat individuals as autonomous agents but also tries to provide protection for people with diminished autonomy. The latter type of human subjects are seen as members of *vulnerable populations*—people who may be at risk of experiencing undue influence or coercion.

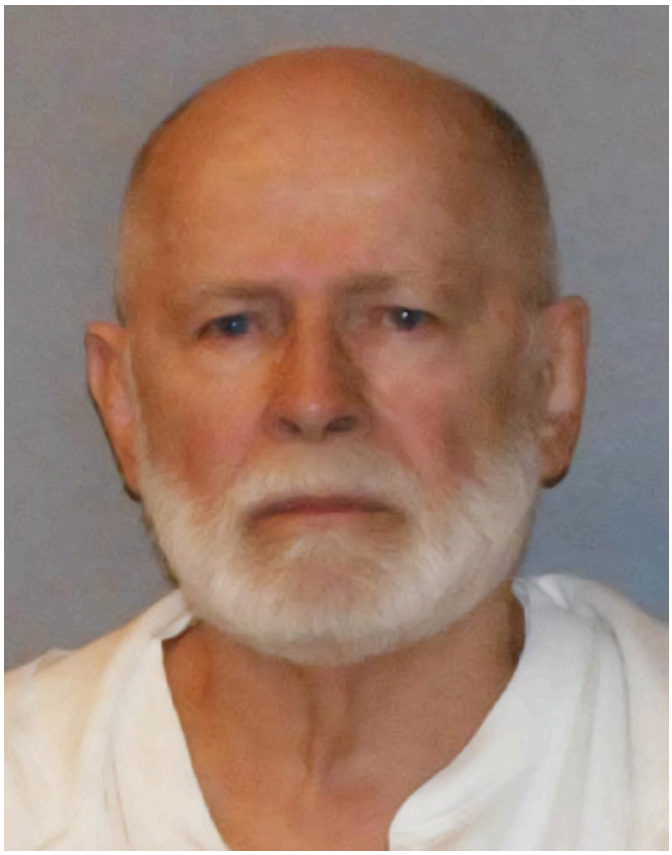


A mugshot of James "Whitey" Bulger from 1956. When Bulger was serving a sentence in a federal prison, the future mob boss and fugitive volunteered for a Project MKUltra experiment on LSD, which researchers said was intended to find a cure for schizophrenia, but which was secretly related to the CIA's studies of mind control. *USP Atlanta, via Wikimedia Commons*

Minors and people who are in jail or prison are among these vulnerable groups. For instance, the incarcerated may feel compelled to participate in research in the belief that they will receive a reduction in their sentences. Another reason that IRBs look closely at research involving prisoners is the grim historical record of how scientists have callously exploited this population for their purposes. For instance, as part of the Project MKUltra experiments discussed earlier, CIA scientists drew many of their human subjects from American prisons and detention centers in Japan, Germany, and the Philippines. Even when individuals volunteered for MKUltra's experiments, they were often deceived about the true nature of the research. The mob boss James "Whitey" Bulger—one of the FBI's most wanted fugitives before his capture in 2011 after 17 years in hiding—was in federal prison on bank robbery charges when he volunteered to be part of an study that scientists said was intended to find a cure for schizophrenia. In reality, the CIA was examining the long-term effects of LSD on the brain as part of its interest in mind control (Gross 2019). Bulger and other inmates signed up for the study in exchange for lighter sentences, but over the course of taking LSD every day for more than a year, Bulger deeply regretted that decision. He began questioning his sanity and contemplated suicide, Bulger later recounted, and he felt particular rage toward the doctor running the experiment, whom he called "a modern-day Dr. Mengele"—a reference to the sadistic Nazi doctor who experimented on inmates at the Auschwitz concentration camp (Curran 2011).

For vulnerable populations, the rules for seeking consent are understandably more stringent. Minors must have the

explicit consent of a parent or legal guardian to participate in research, and IRBs typically require the children themselves to understand and agree to (as much as possible at their given age) the purpose and procedures of the study, as well as its potential risks and benefits. Researchers may be required to ask younger children to provide verbal assent in place of a signature. They may also need to structure the consent process in such a way as to minimize any pressures that parents or guardians may place on children to cooperate. As for research on prisoners, IRBs generally look dimly on such studies, given the inherent inequality in power between those incarcerated and those who might study them.



A mugshot of Bulger from 2011, when he was finally captured. *U.S. Marshals Service, via Wikimedia Commons*

Protection of Identities: Anonymity and Confidentiality



Journalist Ronald Kessler speaks with one of the most famous confidential informants of all time, Mark Felt (1913–2008), who in 2005 revealed that he was the "Deep Throat" source who helped break the Watergate story and bring about the resignation of U.S. president Richard Nixon in 1974. *Ronald Kessler, via Wikimedia Commons*

Another important principle of ethical research is protecting the identities of those participating in the study. At the most basic level, participants have a right to know what information about them is or is not shared with others, and these particulars should be spelled out to them during the informed consent process described earlier. It may also be the case that not sufficiently concealing the identities of research subjects endangers their privacy and their public reputations. For instance, if reading the study allows someone to piece together who a specific participant is—which can be very likely when sociologists study a group of people who already know each other—that participant might feel discomfort and embarrassment and could even be in danger of being manipulated through the use of that information. Furthermore, participants who talk about stigmatized behaviors, such as mental health problems, substance abuse, or criminal activity, could be singled out by employers (who routinely do web searches to vet job applicants as well as current employees) or even law enforcement (who sometimes decide to prosecute individuals who admit to crimes in writing or other media). Even beyond the ethical principle of treating participants with respect and protecting them from harm, researchers have practical reasons for promising to safeguard identities. Without such assurances, interview

respondents may not feel comfortable speaking honestly about their views or experiences, in the fear that others will criticize, shun, or retaliate against them.

To address these potential problems, researchers often promise to maintain either the **anonymity** or the **confidentiality** of their research subjects. Anonymity means that *not even the researcher conducting the study* should be able to connect specific pieces of information provided by participants with their actual identities. An example of anonymity in scientific research is an internet survey in which no identification numbers are used to track who is responding to the survey and who is not.

Although anonymity is the most effective way to avoid any breaches of privacy in research, it can be impractical to guarantee. For instance, sociologists who use ethnographic observation and face-to-face interviewing get to know about participants in person as part of their interactions. It can be difficult or even impossible in these situations to keep participants' identities hidden from the researchers.

As a result, sociologists often promise *confidentiality* rather than anonymity. Under this arrangement, participants allow the researcher to have access to identifying details in the data being collected, but only the researcher should be able to link specific participants to their stored data. In other words, the researcher can identify the specific person who said a particular quote or did a particular thing described in the research report, but they promise not to divulge that person's identity in any published work or public forum.

A confidentiality agreement implies that a participant's personal information is secure, but it is important to understand that these guarantees are not ironclad—as much as a well-intentioned researcher might wish them to be so. For one thing, depending on applicable laws, researchers who learn that someone (including the participant) is at "high or imminent risk of harm" may need to contact a health professional or the authorities; the same is true if they learn about the possible abuse of a child (Kodama Muscente 2022). These exceptions need to be mentioned in consent forms so that participants are not given a false sense of security.

Generally speaking, authorities can use court orders to compel researchers to release or discuss the data they collect, which means they may be unable to protect the identities of any participants suspected of involvement in activities under legal scrutiny. Conversations between social science researchers and their human subjects do not automatically enjoy the "privileged communication" status that those between health professionals and patients or between lawyers and clients do. Nor are researchers protected by so-called shield laws, legislation enacted in many states that allows journalists to keep their sources and unpublished notes confidential. In 2003, the U.S. Department of Health and Human Services began allowing researchers to apply for **certificates of confidentiality**, which prevent them from having to divulge the identities of their human subjects in any legal proceedings, but very few research projects qualify for this level of protection (Office for Human Research Protections 2010). Nevertheless, the code of ethics for the American Sociological Association (ASA)—discussed further in the sidebar **Professional Codes of Ethics**—is clear that sociologists who promise confidentiality should not cave into that pressure even in the face of legal action: "Confidential information provided by research participants should be treated as such by sociologists even if there is no legal protection or privilege to do so" (American Sociological Association 2018:10).

Researchers have been imprisoned for failing to disclose the identities of their participants. In 1993, sociologist Rik Scarce refused to testify before a federal grand jury investigating an act of vandalism. Then a graduate student at Washington State University, Scarce had been studying radical animal rights activism. Authorities were trying to identify individuals who had broken into research facilities at his university and freed animals used in experiments. They believed Scarce had met the perpetrators and demanded that he discuss his conversations with specific members of the Animal Liberation Front (Scarce n.d.). Citing his moral obligations as a sociologist, Scarce told the court that he would not testify about his conversations with activists, and only agreed—under duress—to disclose nonconfidential details not related to his research. He spent 159 days in jail under a contempt order before a judge finally released him (American Sociological Association n.d.).



Sociologist Rik Scarce spent 159 days in jail after refusing to testify about his interactions with Animal Liberation Front activists (pictured here at a protest in Israel). *Roee Shpernik, via Wikimedia Commons*

Compared with anonymity, confidentiality is a weaker form of protection of identities, and even these agreements are not always possible or desirable for particular kinds of research. For instance, **focus groups** that involve interviews with multiple people at the same time (discussed in **Chapter 10: In-Depth Interviews**) cannot ensure confidentiality because the individuals involved will inevitably know at least some identifying details about their fellow respondents. Furthermore, some sociologists have argued that not using the real names of their participants in their published work can have negative repercussions that may outweigh the benefits of protecting privacy (Duneier 1999; Chen and Goldstein 2024). As journalists have long maintained, having individuals provide their quotes on the record with their names attached creates accountability, making it less likely that sources will be able to lie or manipulate without consequences. Some participants may also prefer to see their names in the finished work, perhaps because they want the public recognition or feel their experiences are silenced or trivialized by the use of a pseudonym. Acknowledging these concerns, some IRBs will allow researchers to follow this more journalistic practice of identifying respondents if the risks of the research are minimal and the persons involved explicitly request the use of their real names—sometimes, by signing a "release of information" form distinct from the consent form.

The consent form itself should provide all the key details about how the participant's identity will be protected, including whether the researchers promise anonymity or confidentiality and how they will go about ensuring those conditions. For instance, when participant data is confidential, researchers will typically assign IDs or pseudonyms to their interview

respondents and then keep a password-protected and encrypted document that has the real names associated with those IDs or pseudonyms. The consent form should describe what security measures are in place and who has access to the secure files. If the participants are being taped or video-recorded, the consent form should note that fact and clearly state what will happen afterward to those recordings, which by their very nature include a wealth of identifying details. For instance, recordings may be transcribed afterward and then destroyed; the transcripts may be stripped of any identifiers, such as mentions of the names of the respondents or their family and friends, adding a further layer of protection.

Whether the identities of a study's participants are anonymous or confidential, it may still be possible for an individual who personally knows that person to spot them in the published paper—perhaps based on the configuration of details provided about their age, gender, race, or education. Qualitative researchers may try not to provide too much in the way of detail about a particular individual in order to minimize this risk. When quantitative researchers release their datasets for public consumption, they will sometimes provide the data in an aggregate form that conceals any microdata—the responses of specific individuals—so that people's privacy is protected. For similar reasons, government agencies and polling organizations routinely restrict access to data collected from small geographical units, such as zip codes or small cities, given the risk that—within such a narrow pool of respondents—someone could conceivably connect a person's survey answers to an actual identity.

Sometimes, researchers go so far as to change details—usually minor ones—in order to make individuals unrecognizable in published work even to people who know them. Again, the field of journalism has a different moral take on this practice. (Victor, a former journalist, writes about these divergent sets of professional ethics in the sidebar **The Journalist and the Ethnographer**.) Most journalists would probably be aghast at even slight tweaks to facts within a write-up. In their view, it would cross the ethical line that separates omission (justifiable in some contexts, such as when reporters omit the names of whistleblowers who fear retaliation) from falsification (inherently wrong).

It's also worth noting that the ethical standards for protecting privacy in the digital age are evolving and the subject of much confusion among researchers. After all, a vast amount of data is now available about almost every individual across the online spaces they frequent. Sociologists wrestle with how much of the data—including freely shared social media posts by people who are not public figures—can be used for research purposes without violating reasonable expectations of privacy. For their part, IRBs are not always in agreement about what procedures should be in place to protect the creators of publicly shared digital content, much less data retrieved from less freely accessible spaces online.

The Journalist and the Ethnographer, by Victor Tan Chen

I'm a sociologist now at Virginia Commonwealth University, but I used to be a newspaper reporter (at *New York Newsday*), and as labor of love I still edit a magazine called *In The Fray*, a publication devoted to personal stories on global issues.

When I had aspirations to be the next Bob Woodward back in college, I remember stumbling upon *The Journalist and the Murderer* (1990), a book by *New Yorker* writer Janet Malcolm. It begins with an incendiary paragraph:

> Every journalist who is not too stupid or too full of himself to notice what is going on knows that what he does is morally indefensible. He is a kind of confidence man, preying on people's vanity, ignorance or loneliness, gaining their trust and betraying them without remorse. Like the credulous widow who

> wakes up one day to find the charming young man and all her savings gone, so the consenting subject of a piece of nonfiction learns—when the article or book appears—his hard lesson. Journalists justify their treachery in various ways according to their temperaments. The more pompous talk about freedom of speech and "the public's right to know"; the least talented talk about Art; the seemliest murmur about earning a living. (Malcolm 1990:3)

The Journalist and the Murderer is an account of the relationship between bestselling journalist Joe McGinniss and the subject of one of his true-crime books, physician Jeffrey R. MacDonald. During the course of McGinniss's research for the book, MacDonald was tried and convicted of the murders of his wife and two children. *The Journalist and the Murder* excoriated McGinniss for allegedly "conning" his subject—first by befriending him, and then betraying that confidence.

The unethical behaviors that Malcolm describes in her book are extreme, but they speak to an aspect of journalism that many people find troubling: the way that it uses and manipulates its subjects and then casts them aside, all in pursuit of a sensationalistic headline. This sort of behavior may account in part for why journalists rank abysmally low in Gallup polling on honesty and ethics across various professions. It's part of the reason I decided to go to grad school: I love journalism and believe it plays a vital role in our democracy, but I got tired of the ambulance chasing and other less-than-savory things you sometimes have to do.

IRBs and the profession's code of ethics help sociologists avoid these sorts of problems by setting up protections for the people we interview and observe. This often includes the promise of confidentiality, which can shield our respondents from the public humiliation or retribution at times endured by the subjects of news articles after publication.



Journalists are known for their adversarial interviews, which seek to hold powerful people to account by posing tough questions and pushing back on false claims. Sociologists tend not to pursue this approach, instead stressing empathy and nonjudgmental listening in their in-depth interviews. Axios, *via* Know Your Meme

Before we pat ourselves on the back, however, sociologists do still run into problems at times in terms of how we present our research to respondents and how they ultimately respond to our work. As someone who teaches research methods, I particularly like Jonathan Rieder's *Canarsie* (1987) and Annette Lareau's *Unequal Childhoods* (2011) as examples of how sociologists have dealt with this difficult ethical terrain—particularly the appendix to Lareau's book, in which she describes candidly and thoughtfully the hostile reactions some of her respondents had to their portrayal in her book, in spite of the fact that she hid their identities.

Like journalists, can we also be confidence men and women—gaining trust and betraying it? Furthermore, do we have to do that to gain access to begin with, and to be truthful to the reality we describe? That's the age-old question in research ethics, of course.

Interestingly, journalists would say we are guilty of the exact opposite professional sin: being "overprotective" of our respondents. The fact that their identities—and sometimes those of the cities and organizations we research, too—are hidden leads to a number of complications. First, it's hard to prove to people—particularly skeptical journalists—that what we've written is true. What's to stop us from fabricating our data whole cloth? One obvious safeguard would be the peer-review process—and yet it's not hard to imagine how a determined fabulist could get around even that hurdle.

An ethnographer needs to be cautious in making claims because of the inability in many cases to prove that what you wrote is, without a doubt, true. Of course, in part that's not even up to you, because of the ethical/IRB need or norm of protecting respondent identities. However, I do think one of the strengths of ethnography is its ability to stumble across unexpected situations or outcomes, which in turn can help refine or challenge our theories (with all the caveats that the sample is almost always small and unrepresentative, etc.). But those findings will naturally lead to skepticism because they don't fit with people's preconceptions—and, if they're unflattering to certain people or groups, they may also lead to vicious pushback, however unwarranted it is.

Fact-checking helps journalists to avoid this problem. I've worked as a fact-checker before: what typically happens is the reporter gives you the contact info for their sources, and you call them and verify each quote and fact. It's harder to make up stuff when someone is looking over your shoulder in this way. Even when there's no outright fabrication involved, we as sociologists can alienate readers with our methods. We care about protecting our subjects to the point that in our published work we change (hopefully inconsequential) details, create composite characters, and otherwise alter the reality that we actually observed. For some readers, this is a no-no. Consider, for instance, the outcry over the revelation that James Frey changed or fabricated details in his 2004 memoir *A Million Little Pieces* (and this was a memoir—a genre of literature that has long had a tradition of embellishing the past).

Of course, disgraced journalists like Jayson Blair and Stephen Glass remind us that journalism has failed to catch many acts of dishonesty—and with today's news budgets so strapped, publications no longer have as many resources to verify the information in each article. To be honest, daily journalism operates routinely with a pretty low standard of verifiability. Yes, sources are increasingly recorded on tape or even video, providing documentary evidence, but much of the time reporters are just writing things down in their spiral notebooks. They simply don't have the time to do much else, given deadline constraints. Also, recording an interview changes the dynamic—encouraging the source to use her bland "on the record" voice—and journalists don't necessarily want that. But not taping a source means that they might quote people who then go on to say they were misquoted, and it becomes a he-said-she-said situation. (That happened to me once: a low-level government official made an off-the-cuff comment that he later regretted, and afterward started telling people I made up the quote. I called him and chewed him out for doing that, but there was no way for me to "prove" to other people he had lied because I hadn't recorded him.) This is something that happens more often than you'd think, and that's because journalists (like ethnographers) are dealing with messy real-world constraints.

As someone who has experience interacting with journalists, I know they also look with great skepticism at "anonymous" sources. As they see it, stories based on information collected in this way are by their very nature untrustworthy. Journalistic norms (and sometimes a publication's own policies) emphasize that there has to be a powerfully compelling reason to grant someone anonymity in an article. Now, it's also very true that many people need a promise of confidentiality to feel comfortable telling their story completely and truthfully. And it goes without saying that sources—even nonelites—will exploit the fact that their real names are being used to profit from the attention in some way. For example, when I was working as a reporter, at times I had the hunch that someone was telling me a sob story to garner sympathy and get donations from the newspaper's readers.

If sociologists would say interviewees are more willing to be candid about their personal lives and personally held beliefs if they have the protection of a pseudonym, journalists would stress the fact that hiding their identities can also encourage them to lie: no one can come after them for making up a damaging story about someone else, for instance. And respondents themselves sometimes don't want confidentiality. They may be disappointed to learn their real names won't be published. When I was working as a journalist, I found that people would divulge sensitive details to me or other reporters—for example, about some trauma they'd experienced—and afterward they would tell us they were happy to see their name in print. It gave them a sense of validation to see their story out there and have other people know they actually experienced this. Sometimes, they were contacted afterward by people who related to their story or wanted to help them, and they said they were grateful for that opportunity.

In short, if journalism opens itself up to pernicious forms of exploitation—what I think Janet Malcolm was getting at—in terms of using people and not considering more carefully the consequences of quoting them in a story, the use of anonymity by sociologists can also make it easy to twist our respondents' words and otherwise exploit them. So it seems both fields have their own Achilles' heels, and perhaps we just need to accept they go about things in different ways that are morally justified on their own terms.

It's important to recognize the various ethical and practical trade-offs of all these approaches—not just the distinct practices of journalism and ethnography, but also the different ones used within each tradition. After all, you will find many different approaches to the protection of identities among the most ethical of ethnographers—all of whom, let's stipulate, are trying to do right by both their respondents and their research. Some people just use pseudonyms, some people change details (but only a little), and some people go all out and create composite characters. I can see the ethical rationale for all these approaches. (And in any case, I can't imagine a room full of ethnographers could be forced to pick any one strategy as the professional best practice, even under pain of death.)

To the extent that sociology wants to be part of public debates on important issues, though, I do feel that skepticism about our data is a key reason for lay readers to dismiss our work. In part, this is because readers just don't understand the reasons that we believe practices like confidentiality are so important—they're used to how journalists do their job. But I can imagine they'd have problems even if they understood our reasoning. Why should they trust us? Especially on controversial topics, how do they know we're not lying, or at least fudging the facts? It's not just the question of honesty; it's also a question of style. Using pseudonyms comes across as a bit hokey—especially for place names, which I imagine sound like the egghead equivalent of "Gotham" or "Metropolis" to nonsociologists.

I'm not sure how to deal with these problems. I do think it'd be helpful if sociologists read more journalism (and journalists more sociology) and learn from some of the best practices of the other approach. For sociologists, reading classic works of journalism—from James Agee's *Let Us Now Praise Famous Men* (1941) to H. G. Bissinger's *Friday Night Lights* (1991)—can be incredibly illuminating. It can allow us to draw from the literary beauty, perceptiveness, and heft of these writers in ways that serve our ideas. It can inspire us to write without jargon, make our theories more intelligible to lay readers, and not be afraid to reveal to readers the emotional power of our narratives. Those are the best ways, I think, that we can ensure sociology gets read by the people who could best benefit from its messages.

This sidebar was adapted from two posts that originally appeared on the sociology blog orgtheory.net.

Key Takeaways

1. The *Belmont Report*, a key government document that provided a legal framework for research ethics in the United States, specified three principles for the ethical conduct of research: respect for persons, beneficence, and justice.
2. One way that sociologists can avoid harming research participants is by protecting their identities through practices of confidentiality and anonymity.
3. Sociologists should refrain from misleading participants. They should provide adequate information about their study through procedures that ensure informed consent.

Exercises

1. Let's say a college professor has asked her students to participate in a research study she is conducting. In the syllabus of the course, the professor clearly informs them that they are required to take part in the study. Has the professor obtained her students' informed consent in an ethical manner?
2. Let's say another professor informs his students that they are not required to take part in a research study he is conducting, but those who agree to participate will be given extra credit. Has the professor

obtained his students' informed consent in an ethical manner?

3. A group of researchers are conducting a study in a lower-income neighborhood. They tell the potential participants that the study is on "social issues" and that they will be given gift cards if they agree to take part. Has the research team obtained the potential participants' informed consent in an ethical manner?

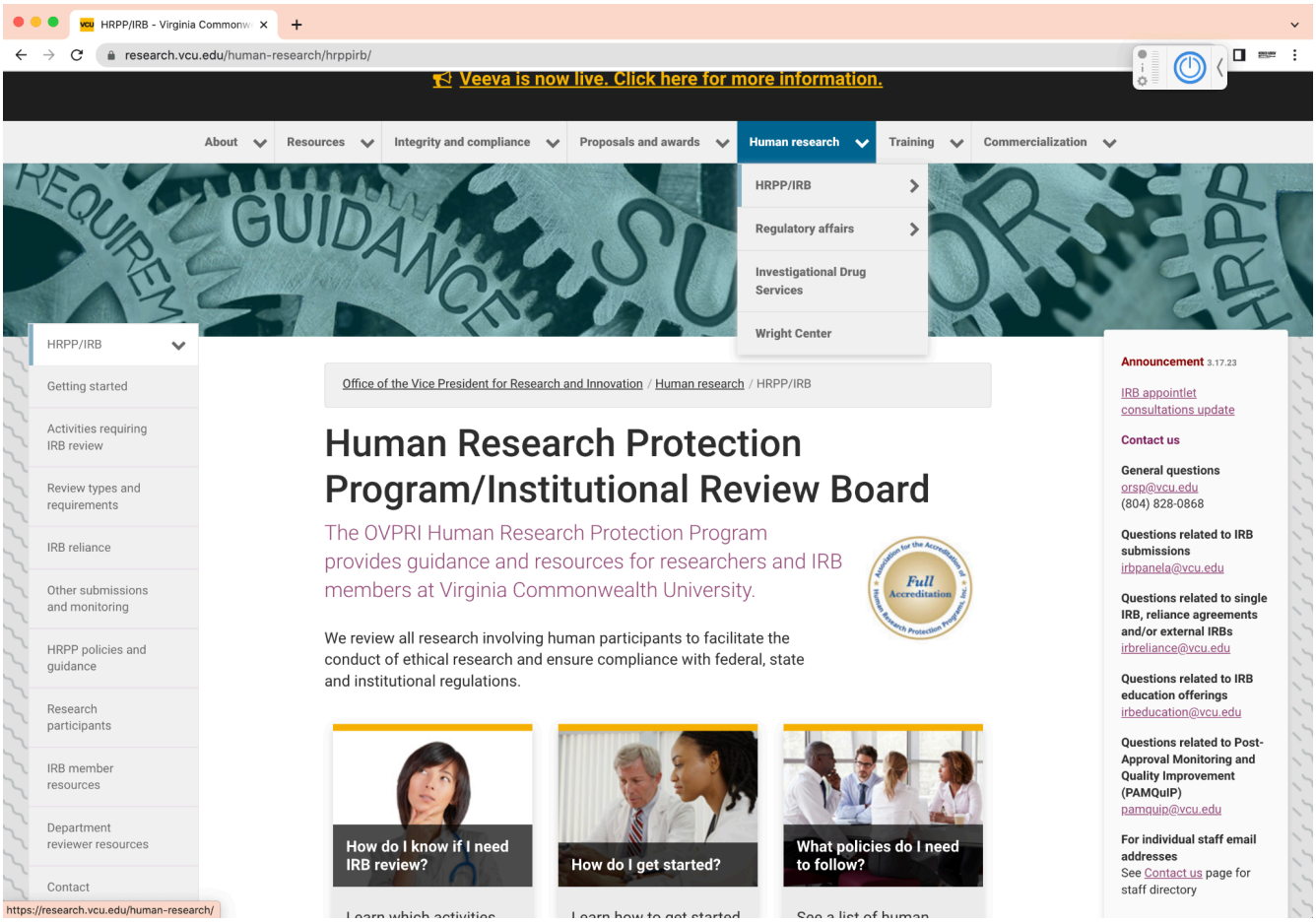
8.3. Regulating Research: Institutional Review Boards and Professional Norms

Learning Objectives

1. Describe the purpose of IRBs and provide examples of the ways they provide oversight over scientific work.
2. Discuss how professional ethical norms influence the conduct of research.

Now that you are familiar with some of the ethical principles and standards that sociologists need to follow in conducting sound research, let's talk in greater detail about the institutional bodies that help devise and enforce these norms and rules. In the United States and other countries, universities and other institutions that want to apply for government funding to support their research projects must establish that they have complied with laws governing the rights and protection of human subjects. At universities, this compliance process is overseen by a panel of experts—called an *institutional review board* in the United States and an *ethics committee* in the United Kingdom.

Seeking approval from IRBs to conduct research is standard practice across scientific communities. Technically, IRBs are required only for institutions that use federal funds, but today, they exist at virtually all scholarly organizations that conduct research. Their purpose is to evaluate the ethics of the studies that the organization's researchers conduct. Especially if their proposed study involves human subjects, researchers must submit an application to the IRB that details their study's design and procedures. The IRB reviews these applications to ensure that the researchers adhere to the principles we described earlier of no harm to participants, informed consent, and protection of identities.



The website of the IRB at our university, Virginia Commonwealth University (VCU) in Richmond. *VCU*

Even though IRBs were originally created to comply with federal laws regulating federally funded projects, they typically vet all research involving human subjects being conducted at a particular institution—including studies not funded by grants, as well as student projects. In fact, the same standards and procedures normally apply to student-conducted research as well, which means that if you are surveying, interviewing, or observing human subjects for a class paper or student thesis, you may be required to go through the IRB review process. You should check with your institution if that is indeed the case. Studies that are not considered "research on human subjects"—say, a statistical analysis of an existing, publicly available dataset, or a content analysis of published news articles—may be designated as exempt from IRB review, which means they can proceed without the review board's approval; the specific procedure depends on the institution, and you should always check what the relevant rules are.

The IRB approval process requires completing an extensive application that covers how the research will be conducted and how the rights and well-being of any human subjects will be protected. In addition, the researchers typically must submit drafts of any consent forms, survey questionnaires, and interview guides they are using. They must take a research ethics course or refresher (usually in the form of self-paced online modules) to certify that they are familiar with the principles of ethical research we have covered in this chapter. And they must identify all the research personnel involved with the study, including the key researchers (called **principal investigators**, or PIs) and any assisting researchers (*co-PIs* or *co-investigators*), sometimes also providing their credentials.

Data collection can commence only after the IRB clears the project. In vetting each application, the IRB's reviewers will suggest changes to the study design to minimize any risks posed to human subjects, ensure informed consent, and protect the privacy of participants. The nature of their evaluation can run the gamut from outright rejection of a problematic research design to minor feedback on the wording of consent forms, interview guides, and survey instruments. At some institutions, researchers frequently revise their IRB applications multiple times, continually adjusting their study procedures until the review board is fully satisfied and greenlights the project.

Although IRB review is widely implemented at U.S. universities, hospitals, and even think tanks, some researchers criticize it as too bureaucratic and burdensome. Depending on the institution and the complexity of the proposed study, IRB review may take weeks or months, which can be a problem when a study is time-sensitive. And the nature of the review process effectively discourages certain types of research, critics say—specifically, more experimental, critical, and open-ended studies. For qualitative studies that use an inductive logic, it may not be possible to disclose a lot of information to participants beforehand, as the study may change over the course of the fieldwork—especially when researchers are asking open-ended questions of their respondents and refocusing their study to address any surprising answers they hear.

Some scholars argue that IRBs need to strike a better balance between protecting human subjects and encouraging important research to proceed (Kotsis and Chung 2014). As the review system stands now, they say, its desire to limit liability, implement stringent rules, and standardize practices across very different types of research places unnecessary hurdles in the way of researchers. This is especially the case for those who study controversial or sensitive topics, whose work is apt to be slowed down or stifled because a bureaucracy whose core mission is to reduce risks cannot help but be allergic to such research. Likewise, scholars who employ methods that gather heaps of identifying data (again, qualitative approaches tend to fall in this camp) are more likely to run afoul of the IRB's mandate. The end result of these institutional incentives is a severe constriction of the freedom and creativity of researchers. Some critics go so far as to claim that IRBs are dysfunctional organizations by their very nature, structured in a way that discourages transparency and accountability. And when it comes to the social sciences—where the potential risks to human subjects tend to be substantially lower than for biomedical studies—the powers that IRBs have to regulate this research are way out of step with the actual need to intervene, critics say, meaning that IRBs have become de facto "censors" of social science research (Schneider 2015).

Nevertheless, the long history of scientific misconduct we've described in detail throughout this chapter should remind you how the regulation of research is essential—left to their own devices, scientists continually prioritize advances in their research over the well-being of their research participants. We suggest that you approach IRB reviewers not as antagonists but as collaborators in your research. They are trying to do their job—protect human subjects—and they often have a wealth of knowledge about how scientific studies are conducted and what procedures make sense for a given situation. Ideally, their feedback will strengthen your study and give you confidence you are acting ethically as a researcher.

Professional Codes of Ethics

Lawyers, doctors, and many other professionals can be disbarred or lose their licenses for failing to follow ethical rules. No such provision exists for social science researchers, generally speaking. Nevertheless, the various disciplines have their own sets of guidelines for protecting research subjects and conducting ethical research. Even if there is no established penalty for transgressions, sociologists who deviate from these norms will face criticism from peers and possibly sanction or termination by their employers.



Like many national associations for academic disciplines, the ASA has long had a Code of Ethics, provided on its website, that U.S. sociologists are expected to adhere to as card-carrying members of the profession.

In the field of sociology, the ASA has adopted a set of ethical principles intended to guide researchers in the discipline (American Sociological Association 2018). The ASA Code of Ethics features the following five general principles:

Professional competence: Sociologists should recognize their own limitations and only conduct research for which they have been properly trained. Researchers should also continually update their knowledge and expertise so that they remain competent, including in their understanding of shifting ethical norms.

Integrity: Sociologists should be "honest, fair, and respectful" in all their professional activities—including, but not limited to, their research activities.

Professional and scientific responsibility: Sociologists should show respect to their peers and maintain collegial relationships. At the same time, this principle warns against collegiality if it impedes our ability to behave ethically, conduct our research honestly, and otherwise act in ways that preserve public trust in the soundness of sociological work.

Respect for people's rights, dignity, and diversity: The fourth principle, respect for people's rights, dignity, and diversity, addresses the need to reduce bias in all professional activities and be inclusive in the course of conducting research, in terms of recognizing one's position relative to the individuals and groups being studied and ensuring that a diversity of voices are heard whenever they are relevant.

Social responsibility: Sociologists should "strive to advance the science of sociology and serve the public good."

The ASA Code of Ethics (which you can read in full here) goes into much greater detail about each of these points.

You may have noticed in reviewing these five general principles that they go well beyond the responsibility of just treating research subjects humanely and respectfully—which, as we said, tends to be the focus of the vetting done by IRBs. In the last section of this chapter, we'll touch on other dimensions of the key principle of research "integrity" by discussing the problem of fraudulent and biased research.

Key Takeaways

1. IRBs in the United States (and similar bodies in other countries) ensure the ethical conduct of the research that takes place at universities and other organizations.
2. All U.S. institutions receiving federal support for research must set up IRBs. Even when their researchers do not receive federal support, however, organizations may choose to set up IRBs to oversee studies and ensure that they abide by ethical standards, such as by vetting consent forms, reviewing procedures for recruiting participants, and mandating that researchers securely store any data that contains identifying information.
3. Because of professional codes of ethics, scientists must adhere to standards of integrity and responsibility or face potential backlash from their peers and employers.

8.4. Veracity and Objectivity in Research



In the Academy Award–winning documentary *Inside Job* (2010), filmmaker Charles Ferguson interviewed Columbia University economist Glenn Hubbard, the former chair of President Bush's Council of Economic Advisers. During their increasingly heated conversation, Ferguson pressed Hubbard about his failure to disclose the compensation he received from various financial firms even as he spoke out publicly against regulating the industry—revelations that ultimately led both Columbia and the American Economic Association to change their policies regarding conflicts of interest. *Sony Pictures Classics*

Learning Objectives

1. Discuss the importance of objectivity in social scientific research.
2. Identify possible political concerns related to research ethics.

As we describe in **Appendix A: Presenting, Writing, and Publishing**, it is critical for researchers to provide a full and detailed accounting of how we did our work (typically described in the methods section of a presentation or paper) and what our analysis concluded. When we disclose all this information, then those who draw on our work—say, to conduct other research, or to formulate and implement policies—can trust it and make the best possible decisions. This transparency also assures readers that we have conducted legitimate research and did not simply come to whatever

conclusions we wanted to find. In the appendix, we talk about these practices as a matter of ensuring the quality of our research and contributing to the scientific conversation: we want fellow sociologists to know all the details so that they can assess our work fairly and fully, and so that others might be able to easily replicate or build on our work. But exhaustively providing details about the research we do is also a moral issue, one that has, sadly, continued to crop up in public debates over the truthfulness of scientific research.

Full transparency in research entails being completely honest about how we identified and recruited our research subjects, how we collected and analyzed our data, what the specific limitations of our methodological approaches were, and what our full and unvarnished findings were—even when they contradict what we expected or hoped, or what prominent scholars have concluded. Furthermore, we as researchers should be ready to share any data we collected (properly anonymized to protect people's identities, of course) with the larger scientific community. Sometimes researchers may leave out methodological details because there may not be enough time or space to share them. This is often the case with news reports of research findings, which (as we also discuss in the appendix) usually must be condensed and simplified.

Researchers may also have less justifiable reasons to conceal. To prevent others from challenging their work, they may be silent about the weaker parts of their studies, or hide or fudge findings they may not like. They may refuse to share their data—perhaps in the hopes that they can crank out as many scholarly publications as possible before other researchers can catch up. And they may jealously guard the details of how they did their work—because they fear copycats, perhaps, or because they want to hide flaws in their studies. Yet science relies on transparency to work its magic of generating verified knowledge and promoting creative and critical thinking. Good researchers are upfront about their work—or at least as upfront as they can possibly be, given other ethical obligations like protecting the identities of their participants. And as we've stressed previously, every research study—especially any ambitious one—has flaws and shortcomings. We, as serious scholars, have to own up to them. In fact, if you find that other researchers are less than open about their methods, data, and findings, you might want to question whether you should believe what they say.

As scientists, we have an obligation to uphold veracity in our research—put simply, to tell the truth about whatever we find. But for professional and even financial gain, sometimes scientists do not fulfill these obligations. There are many examples to pick from, but one relatively recent and egregious example is the fraudulent research conducted by Harvard evolutionary biologist Marc Hauser. An early sign of problems with Hauser's research was a 1995 study he led, published in the prestigious *Proceedings of the National Academy of Sciences*, which concluded that cotton-top tamarins could recognize themselves in mirrors. Psychologist Gordon Gallup reviewed some of the available videos of Hauser's experiment and saw "not a thread of compelling evidence" of such behavior among the tamarins featured in the recordings (Johnson 2010). When he asked for the remaining tapes, he was told they had been stolen (*Economist* 2010). In response to the critique by Gallup, Hauser first defended the study and clarified the criteria used to code the video data; in 2001, however, he said his subsequent work had failed to replicate the earlier finding that tamarins recognized themselves in mirrors.

Rumors circulated for years about Hauser's dubious research, and in 2007, Harvard University opened up an investigation, ultimately finding him responsible for eight counts of scientific misconduct. Although details of the internal investigation remain secret, individuals who worked with Hauser have alleged that Hauser falsely coded video data of monkey behavior—even pressuring research assistants who wanted to accurately recode the data not to do so. In the face of these criticisms, Hauser resigned from Harvard in 2011, and a year later, the Office of Research Integrity, a U.S. federal agency overseeing researcher ethics, found him guilty of scientific misconduct.

Fraudulent research can cause harm that goes well beyond misled and irate colleagues. This is especially the case nowadays, when scientific findings are regularly attacked and dismissed. During the early years of the Covid-19 pandemic, many people refused to take the various vaccines that were quickly developed for the virus and its variants. Hesitation about any new technology is understandable, and every medical treatment has risks, but the international

scientific community was flabbergasted by the extent to which nonscientists challenged their work, or latched onto unproven treatments for Covid that, at best, had support from small-*n*, one-off studies. Vaccines in general had a very safe track record, and studies of Covid vaccines had found only small risks associated with them, especially relative to the acute and chronic conditions threatened by Covid. The incendiary public debates over Covid vaccines and treatments often drew from anecdotal evidence, misrepresentations of scientific studies, or outright false claims. And while opposition to Covid vaccines in the United States became closely associated with conservative politics, it built on a broader current of skepticism about vaccines that had thrived in the decades before Covid. Across the country, parents' decisions not to vaccinate their children based on unfounded fears have regularly led to outbreaks of infectious diseases like measles, which in 2000 had been temporarily eliminated (Centers for Disease Control and Prevention 2020).

There is only so much that scientists can do in the face of **misinformation** (inaccurate information) or **disinformation** (deliberately misleading information), but at the very least, they can present research—whether or not it is their own, and whether or not it supports the findings of their past work—in an impartial and fair fashion. They can also help nonscientists understand how to evaluate research and think critically about the claims being made (Dyer and Hall 2019). Unfortunately, well-publicized scandals in science have eroded public trust and fed contrarian views with no real empirical support. The vaccine debate provides another relevant example. A 1998 paper published in the *Lancet*, a top-ranked British medical journal, made the claim that the MMR vaccine (which combines vaccines against measles, mumps, and rubella) was linked to the development of autism in children. Other studies failed to replicate this finding, however.

The reporting of investigative journalists eventually uncovered evidence that the lead researcher on this study, physician Andrew Wakefield, had fudged diagnoses and dates to fabricate a causal relationship between the vaccine and autism. Wakefield reportedly had been set to make millions from selling a diagnostic kit and a competing vaccine that would benefit from skepticism about the MMR vaccine. The *Lancet* and Wakefield's coauthors retracted the paper, and Wakefield was discredited and stripped of his license to practice medicine. Although research in subsequent decades has continued to find no relationship between autism and the MMR vaccine—or any other vaccine—the *Lancet* study appears to have fueled skepticism about vaccines that continues to this day.

The scandal over the retracted *Lancet* study also underscores problems with bias in research. In Wakefield's case, he clearly had a **conflict of interest—**a situation where a researcher would derive a financial or other personal benefit from a particular research finding, and therefore might favor such a finding (consciously or unconsciously) while conducting their research. In Wakefield's case, he was studying whether or not the MMR combination vaccine contributed to the development of autism, yet finding a causal relationship would have brought him substantial financial gain—meaning his objectivity as a researcher was in question, and it would be hard to trust the conclusions of his study.

Typically, reputable journals and publishers require authors to disclose any conflicts of interest. While Wakefield's conflict of interest was particularly sensationalistic, more mundane but still troubling conflicts might include receiving grant funding from an organization that favors the policy, program, or intervention you are studying, or conducting research that would benefit a family member's financial or career stake in regard to the issue being studied. To track potential conflicts of interest, IRBs sometimes ask researchers to regularly report their financial interests (such as their holding of stock in certain companies), which are recorded officially and can be used to identify potential conflicts. When such situations actually do arise, a journal's editors may decide not to publish a scientist's work, or they may publish it with a statement (typically at the end of a journal article) that fully discloses the existence of that conflict of interest. In the latter case, an editor might think the researcher can still conduct the research in an impartial fashion even though the conflict exists. Yet this determination is always a subjective judgment call, and to some extent, the reader will need to decide how large of a grain of salt should accompany their take on the study's results.

Conflicts of interests are exceedingly common in the social sciences, given that the work we do often has real-world consequences for policymaking. In fact, in the aftermath of the 2008 financial crisis that tanked the global economy, many economists got into trouble for failing to disclose the financial interests they had even as they commented on issues relevant to those interests. Some economists, for instance, produced papers and reports before the crash that

dismissed some of the worrying signs of the economy's fragility and opposed efforts to regulate the financial industries whose fraudulent and reckless practices ultimately brought about disaster. In the documentary *Inside Job* (2010), an Academy Award-winning analysis of the factors that brought about the financial crisis, filmmaker Charles Ferguson famously criticized a number of these economists, including scholars working for both Democratic and Republican administrations (Inman and Kingsley 2011). For example, in a contentious interview with Glenn Hubbard, former chair of President George W. Bush's Council of Economic Advisers, Ferguson confronted the Columbia University economist about his strong advocacy for deregulating financial markets while also being paid hundreds of thousands of dollars by financial firms for serving on their boards or consulting for them—positions that were sometimes disclosed on his website, but not mentioned in his writings (Solman 2011). Afterward, Columbia strengthened its conflict-of-interest policies, and the American Economic Association, the professional association representing the country's economists, adopted stronger standards for such disclosure in its code of ethics (Lahart 2011).

Bias in research does not have to be deliberate or devious. People are not always aware of how much their personal interests or background can influence their take on things, including their take on their own research: remember the power of **confirmation bias** in skewing our findings. For example, pioneering American and European anthropologists who immersed themselves in fieldwork in remote indigenous communities produced powerful accounts of life in small and close-knit villages without many modern technologies. Yet even luminaries like Margaret Mead—who in her classic ethnography *Coming of Age in Samoa* (1928) described Samoan youth as having sex in a free and open manner—have been accused of "romanticizing" or misrepresenting aspects of another culture, arguably because their Western perspective shaped how they interpreted what they observed.

Social scientists, particularly those who use qualitative methods, are more skeptical today about outsiders studying cultures they are not part of and do not fully understand. They worry that ethnographers do not always fully appreciate the impact of their work on those they study or the voyeuristic way they can approach less advantaged communities—for instance, how they personally profit through their academic publications from the "deviant" and "exotic" behaviors they observe in low-income neighborhoods. On the flip side, ethnographers have also been criticized for uncritically embracing the viewpoints of the communities they study—which, at an extreme, can reach the point of abetting immoral and even illegal activities. (We will discuss these various pitfalls at greater length in **Chapter 9: Ethnography**.) Columbia University sociologist Sudhir Venkatesh, who penned the blockbuster sociology trade book *Gang Leader for a Day* (2008), was subsequently criticized for contributing to the criminal enterprise of the drug gang he studied. Separately, Venkatesh came under fire for being careless with facts in his writing and not being able to account for missing funds at a center he oversaw at Columbia (Kaminer 2012).

Ethnographer Alice Goffman wrote another widely read book, *On the Run* (2015), which followed young African American men as they interacted with police in a low-income neighborhood in Philadelphia. Like Venkatesh, fellow sociologists at first praised her work as providing an unprecedented and empathetic look at individuals on the margins of society as well as the institutional forces that hemmed in their lives. However, Goffman later received criticism for becoming too close to her informants, such as when—according to her own account—she drove an armed informant around as he sought revenge for the murder of another of Goffman's respondents. In addition, some academics questioned the accuracy of police practices described in her book and even alleged she had exaggerated or made up facts about her observations. Journalist Jesse Singal tracked down some of the families and individuals described with pseudonyms in Goffman's work and confirmed they were real (itself an illustration of both the distinct professional norms of journalists and the difficulty of preserving confidentiality for participants). Singal concluded that "her book is, at the very least, mostly true" and that her critics' accusations of embellishment or factual errors were likely due to the fact that Goffman "simply didn't heed her own advice about credulously echoing sources' stories" (Singal 2015b, 2015a).

This brings us back to the divide between the ethical tenets of professional journalism (as much as they are often savaged) and those of sociology. From a principled perspective, journalists see themselves as doggedly pursuing the truth—anyone else's feelings be damned. As rock journalist Lester Bangs advised a cub reporter in the film *Almost Famous*

(2000): "Be honest, and unmerciful." For many journalists, the professional imperative is to hold people in power to account. Therefore, they are more likely to conduct adversarial interviews, challenging people's opinions or narratives of events, laying down verified facts, and catching any lies or contradictions. By contrast, the model in sociology tends to be empathetic interviewing, in which the goal is to listen without judgment, build rapport with the interviewee, and gather as much detailed and in-depth information as the interviewee is willing to provide. Such an orientation toward interviewing fits well with our ethical obligations as sociologists not to exploit or harm participants, treating them with a recognition of our position relative to them and the power we wield in that relationship (see the discussion of manipulation in the sidebar **The Journalist and the Ethnographer**). One downside of this approach is that it may lead sociologists to naively parrot what their respondents tell them, without checking whether their claims really hold up. To manage this risk, it is a good idea to speak to multiple sources and **triangulate** one's findings—a strategy we discuss further in later chapters. In rare cases, sociologists have gone so far as to hire professional fact-checkers to verify the details in their books, as Matthew Desmond did for his Pulitzer Prize-winning book *Evicted* (2016); fact-checking is expensive, however, and therefore out of reach of most researchers.

Of course, as we discussed in **Chapter 3: The Role of Theory in Research**, some social scientists dispute the idea of "truth" altogether, and they push backs against notions like journalistic "objectivity" that can seem to them like a convenient excuse to impose a mainstream or establishment interpretation of the facts. It is also true that, as sociologists, we are also not so much interested in individuals—say, the decisions of one specific leader—but rather in how the actions and thinking of those individuals reflect broader social forces. This places less weight on the "truth" of particular persons or situations, given that they are being used as representations of more abstract cultures or social structures. Sociologists might also pay more attention to who is telling a particular story. For instance, another criticism leveled against Goffman's study of young African American men running from the law was that she is a white woman from a privileged background (in fact, she is the daughter of the influential sociologist Erving Goffman). Was it appropriate for someone like her to tell the story of the men she followed (Brown 2017)? Social scientists often disagree about these thorny ethical issues, and as we've suggested, journalists can legitimately take a very different approach based on the distinct professional norms they operate under. Even if we as sociologists take to heart the postmodernist critiques of "objectivity" and "truth," it is always worth considering how fair our interpretation of reality is, and whether it includes a sufficient range of viewpoints and facts to capture what is going on in a reasonable, open-minded, and valid way.

Politics in Research



Academics are often singled out by authoritarian regimes for murder or imprisonment, such as during Taiwan's "228 Incident" (depicted in this woodcut print *The Terrible Inspection*, by Taiwanese artist Huang Rong-can), when Nationalist soldiers killed an estimated 18,000 to 28,000 people, including much of the island's intellectual elite (Forsythe 2015). *Huang Rong-can, via Wikimedia Commons*

A final point we want to make about ethics is how much the political and cultural context we operate in can shape our ability as scholars to produce incisive, high-quality research. A repressive or polarized political context can have a chilling effect on important research that does not fall in line with popular ideologies.

Especially in countries with limited civil liberties, researchers may feel compelled to legitimize authoritarian practices, reinforce a regime's narratives about politics and history, and even ignore or justify state crimes (Akçam 2007). Not doing so might mean a scholar's ability to conduct research or advance in their career is stymied, given how many universities and research institutes are government-funded. Researchers who are closer to authorities may be given easier access to archives, opportunities to conduct larger surveys among the populace, and financial and other support from state-run institutions. In turn, scholars may omit important findings if they could be perceived as antagonizing those in power or challenging some of the myths the state is trying to propagate. After all, taking such a critical perspective can lead to serious personal consequences: throughout history, authoritarian regimes have targeted academics for killing and imprisonment, seeing intellectuals as inherent threats to the obedience and docile order they desire. It is understandable why researchers who operate in repressive environments will toe the official line, but doing so amounts to a breach of the ethical obligations that social scientists have to be truthful in their research (Öztürk 2018).

Even in countries where scholars presumably have greater freedom to conduct research as they want to, governments regularly try to influence research in more or less innocuous ways. In the United States, scientific funding agencies like the National Science Foundation and the National Institutes of Health shape research agendas with the massive amounts of money they leverage to support specific lines of study. More troubling for many academics is the major role that military agencies play in funding research. For instance, the Stanford Prison Experiment that we described at the beginning of the chapter was funded by the U.S. Office of Naval Research. Henry Murray, the psychologist who subjected the Unabomber and other students to continuous humiliation in his Harvard laboratory, formerly served as a lieutenant colonel in the Office of Strategic Services (often considered to be the precursor to the CIA), where he conducted psychological experiments on mind control and psychedelics; bioethicist Jonathan Moreno (2006) and others have even argued that Murray's Harvard experiments were part of the CIA's classified Project MKUltra.



Military agencies often fund social scientific studies. The infamous Stanford Prison Experiment was sponsored by the U.S. Office of Naval Research, and psychologist Philip Zimbardo first published his research in Naval Research Reviews, the office's journal, which he said was a requirement of the grant. *Stanford Prison Experiment*, *via PrisonExp.org*

So far, we have been talking about how those with political power can silence or suppress the voices of academics. But it is important to note that the field of sociology can also reinforce political ideologies. As we discussed in **Chapter 2: Using Sociology in Everyday Life**, critics have pointed out the ideological conformity in the discipline, with one critic saying that sociology "is so dominated by the left that it is instinctively dismissed by the right" (Kristof 2014). As important as it is to have and hold tight to your own moral and political values, an intellectually honest student of social life quickly realizes how complex it is, and how a variety of different perspectives can be legitimately brought to bear to understand it. Ethically, social scientists should hold themselves to a higher standard than just echoing the acceptable opinions of people on their side of a debate. They should be open to being surprised and persuaded by what they uncover through empirical investigation. If you care about not just preaching to the choir but also convincing others with different perspectives, the most effective strategy may be to conduct your research as honestly and impartially as you can.

Studying Vulnerable Populations and Polarized Issues: A Q&A with Didem Danış



Didem Danış is an associate professor of sociology at Galatasaray University in Istanbul. She is also the chair of the Association for Migration Studies (GAR), a Turkish nongovernmental organization (NGO) she helped found. Born in Istanbul, Danış received her bachelor's and master's degrees at Turkish universities and then completed her doctorate at one of France's top graduate academies—the School for Advanced Studies in the Social Sciences (EHESS), where she wrote her dissertation about the social networks of Iraqi transit migrants in Istanbul. An urban sociologist and scholar of migration, Danış studies how refugees and other migrants assert their agency in a context of structural constraints and deeply rooted historical inequalities. More broadly, she examines how transnationalism, social networks, and gender issues shape immigrant communities. In addition to her research in Turkey, Danış has collaborated on studies of migrants in France and the Middle East and has edited four volumes on topics relating to migration.

Why did you decide to study sociology?

When I was finishing high school, I was very confused. I couldn't decide whether to study archeology, biology, or the social sciences. My grades were high, and I knew I could choose any major I wanted, but most of the people around me were trying to convince me to study business administration. At that time, globalization was at full throttle, and finance had replaced industry. Business administration was a major that seemed to guarantee high-salary white-collar jobs for its graduates.

Yet, I wanted to go against the current. I decided to study political science at Boğaziçi (Bosphorus) University in Istanbul, considered among the best institutions for social sciences in Turkey at that time. Political science was an exciting choice in a country whose political life was interrupted by military coup d'états almost every 10 years. However, at the end of my first year, I realized that it would not be possible to understand the political dynamics without understanding the society, and I decided to do a double major in sociology.

I became fascinated by sociology courses related to cities and the interaction between spatial and social dynamics. After all, I was living in Istanbul, an impressive metropolis and a global crossroads where different migration movements have intersected throughout history. My first readings on urban sociology made me better understand my family's story, which is based on a migration journey, too—one within Turkey. As I read more sociological texts, I started to grasp the inequalities that various migrant groups encounter in the urban space. Istanbul's population has been formed by various immigration waves, yet the city has not embraced every group with the same warmth or offered them equal opportunities.

Global cities are places of discrimination and injustice. But they are also sites of solidarity and struggle. Istanbul—a city of great opportunities and challenges, of opening as well as closing doors—has been my passion as a sociologist.

You have worked closely with national and international NGOs. How has your background as a sociologist influenced your work in the nonprofit sector? Have you found certain research methods to be particularly useful in these professional settings?

As sociologists engaged in the production of scientific knowledge, we have an important role to play in sharing our knowledge about society with the general public. I use the knowledge I have gained through scientific methods to strengthen the work of nonprofit organizations and answer the public's questions about migration, one of the main challenges of our time. Since migration is a global issue, a bridge needs to be built between local, national, and international organizations. As an academic who specializes in the sociology of migration, I try to link these fields.

In environments such as Turkey where quantitative data are limited, inaccessible, or unreliable, qualitative research methods play a very valuable role, providing information that can support the efforts of NGOs and help them develop policy proposals. This is especially the case for migration issues. Informality and irregularity characterize the lives of many migrants. Researchers studying these communities therefore need to be flexible and adaptable—as Stephen Castles (2012:25–26) puts it, willing to "respond to the lessons of the field and to hear what respondents are saying." In the course of research, frequent and unexpected problems arise. Good researchers reexamine and reformulate their original questions as needed.

You study Syrian, Afghan, and Iraqi refugees and immigrants in Turkey. We live in a world that is deeply divided on issues of immigration, with refugee flows in particular sparking anger and violence across countries. What role do you think research can play in helping societies address these demographic shifts in a healthy and constructive way?

In today's public discussions of immigration, we see many features of the "post-truth" era. There are intense campaigns of disinformation and other efforts to manipulate people's thinking about immigrants and refugees, especially on social media. By examining migration dynamics from the perspectives of both immigrants and host societies, sociologists can shed light on processes and experiences that are rendered invisible in everyday political debates. Take "replacement theory," a conspiracy theory that is being peddled all over the world. Replacement theory argues that foreign powers bring immigrants into a country to disrupt its social structure. Over time, immigrants will gain demographic superiority and "destroy our homeland," according to the theory's proponents.

This idea is widespread in immigrant-receiving countries—even though such a drastic demographic shift is just not possible in many of these places. A sociological perspective helps us understand what makes these exclusionary discourses, nevertheless, so popular. People feel that others they consider to be different—minority groups—pose an ideological, social, or economic threat. They thus overstate the size of these groups. This sense of "symbolic threat" is stronger during periods of profound social change. As a sociologist, I try to clarify these dynamics, how they lead to outputs like hate speech, exclusion, and violence toward immigrants, and what risks they pose for society.

You have received vicious attacks on social media because of your work with refugees. Have you been surprised by the backlash you've received? How have you been able to deal with this harsh public scrutiny, and how do you feel it affects your work as a scholar?

Xenophobic and racist rhetoric has become widespread online. Anyone who criticizes these discriminatory discourses and talks about migrants' rights is targeted. Migration researchers get their share of these attacks because they examine the subject with scientific methods and produce evidence-based arguments against disinformation and manipulation. When I first encountered such offensive statements, I was surprised and frightened. Especially on social media, some people use a much more aggressive style—even prying into their targets' private lives. They want to scare and silence scientists who speak out. Eventually, my skin thickened

against such attacks on social media, but what is still difficult for me is to hear the same unfounded claims in face-to-face interactions.

When we sociologists limit our speech to narrow scientific environments, we can comfortably express our ideas. But as someone who believes in public sociology, I think that we should share our knowledge with the wider public. To do so in the face of attacks, we need to use some tools that the sociological imagination provides—for example, drawing strength from online communities of colleagues who adopt similar values. The Facebook group Migration Researchers' Platform, which now has more than 10,000 followers, has been a good support for me in this respect. The researchers involved with this group also founded the Association for Migration Research in Istanbul.

What sorts of ethical challenges have you encountered when you've conducted research on vulnerable populations like immigrants and refugees? Are there particular approaches you take to address those challenges?

A crucial issue in doing research with immigrants and refugees is being aware of the vulnerability and precarity within these communities. Researchers should do no harm—neither to the immigrants, nor to themselves. For me, this means being careful about my respondents' privacy and anonymity and not revealing any confidential information I gathered, which may put them in danger if disclosed. Whenever I am in the field, self-reflexivity and positionality are my key considerations.

You live in a country where academics face intense political pressures regarding how they do their work. What impact does the political situation in Turkey have on you, as a researcher and otherwise?

In my country, freedom of expression has been seriously curtailed over the last ten years. The rise of authoritarianism places a heavy burden of fear on those who produce and disseminate critical knowledge. In particular, authoritarian regimes consider sociology to be a threat, since it is a discipline that studies society in a scientific and analytical way.

Political pressure is not the only danger. As in many other countries, scholars in Turkey are also dealing with the commodification of the academic world. The funding of research has become a mechanism of control and censorship. Both the public and private sector have an instrumentalist interest in the knowledge produced by researchers.

What advice would you give sociology students who want to do work similar to your own—research that helps disadvantaged communities and vulnerable populations?

I recommend that they pursue the subjects that they are sincerely curious about. Sociology has the potential to build bridges between the academic and the real world, and they should keep in mind their capacity to influence politics. As a sociologist, I want to study the society I live in, because I want to change it.

Key Takeaways

1. Even though we sociologists often critique black-and-white notions of truth and objectivity, we can all follow standards of honesty and transparency in our research to ensure that our work is credible and does not contribute to misinformation.

2. Political concerns and authoritarian pressures may hamper scientific research and further complicate notions of what is ethical and appropriate research.

8.5. References

Akçam, Taner. 2007. A *Shameful Act: The Armenian Genocide and the Question of Turkish Responsibility*. New York: Henry Holt.

American Sociological Association. 2018. *Code of Ethics*. Washington, DC: American Sociological Association.

American Sociological Association. n.d. "Case 99. A Real Case Involving the Protection of Confidential Data." Retrieved June 17, 2023 (https://www.asanet.org/case-99-a-real-case-involving-the-protection-of-confidential-data).

Annas, George J., and Michael A. Grodin, eds. 1995. *The Nazi Doctors and the Nuremberg Code: Human Rights in Human Experimentation*. New York: Oxford University Press.

Brown, Sarah. 2017. "Should Alice Goffman's Work Cost Her a Faculty Position?" *Chronicle of Higher Education*, April 26.

Castles, Stephen. 2012. "Methodology and Methods: Conceptual Issues." Pp. 15–36 in *African Migrations Research: Innovative Methods and Methodologies*, edited by M. Berriane and H. D. Haas. Trenton, NJ: Africa World Press.

Centers for Disease Control and Prevention. 2020. "Measles: Answers to Common Questions." *Centers for Disease Control and Prevention*. Retrieved June 21, 2023 (https://www.cdc.gov/measles/about/faqs.html).

Chase, Alston. 2000. "Harvard and the Making of the Unabomber." *The Atlantic*, June 1.

Chen, Victor Tan, and Jesse Goldstein. 2024. "The Theatre of Entrepreneurship: Learning to Perform the Speculative Self in University Entrepreneurship Programs." *Social Problems* 71(1):203–19. doi: 10.1093/socpro/spac008.

Crowe, Cameron, dir. 2000. *Almost Famous*. Dreamworks Pictures, Columbia Pictures, Vinyl Films.

Curran, Kathy. 2011. "Whitey Bulger's Notebook Chronicles LSD Prison Testing." *CBS News Boston* (WBZ), July 7.

Desmond, Matthew. 2016. *Evicted: Poverty and Profit in the American City*. New York: Broadway Books.

Duneier, Mitchell. 1999. *Sidewalk*. 1st ed. New York: Farrar, Straus and Giroux.

Dyer, Kathleen D., and Raymond E. Hall. 2019. "Effect of Critical Thinking Education on Epistemically Unwarranted Beliefs in College Students." *Research in Higher Education* 60(3):293–314. doi: 10.1007/s11162-018-9513-3.

Economist. 2010. "Monkey Business?" *The Economist*, August 26.

Evans, Richard J. 2006. *The Third Reich in Power*. Reprint ed. New York: Penguin Books.

Ferguson, Charles, dir. 2010. *Inside Job*. Sony Pictures Classics, Representational Pictures, Screen Pass Pictures.

Forsythe, Michael. 2015. "Taiwan Turns Light on 1947 Slaughter by Chiang Kai-Shek's Troops." *New York Times*, July 14.

Frey, James. 2004. *A Million Little Pieces*. New York: Random House.

Goffman, Alice. 2015. *On the Run: Fugitive Life in an American City*. Chicago: University of Chicago Press.

Gross, Terry. 2019. "The CIA's Secret Quest For Mind Control: Torture, LSD and a 'Poisoner In Chief.' " *NPR*, September 9.

Humphreys, Laud. 1970. *Tearoom Trade: Impersonal Sex in Public Places*. London: Duckworth.

Humphreys, Laud. 2008. *Tearoom Trade: Impersonal Sex in Public Places*. Enlarged ed. New Brunswick, NJ: AldineTransaction.

Inman, Phillip, and Patrick Kingsley. 2011. "Inside Job: How Bankers Caused the Financial Crisis." *The Guardian*, February 17.

Johnson, Carolyn. 2010. "Author on Leave after Harvard Inquiry." *Boston Globe*, August 10.

Kaminer, Ariel. 2012. "Columbia's Gang Scholar Lives on the Edge." *New York Times*, December 1.

Katz, Ralph V., B. Lee Green, Nancy R. Kressin, S. Stephen Kegeles, Min Qi Wang, Sherman A. James, Stefanie L. Russell, Cristina Claudio, and Jan M. McCallum. 2008. "The Legacy of the Tuskegee Syphilis Study: Assessing Its Impact on Willingness to Participate in Biomedical Studies." *Journal of Health Care for the Poor and Underserved* 19(4):1168–80. doi: 10.1353/hpu.0.0067.

Kinzer, Stephen. 2019. *Poisoner in Chief: Sidney Gottlieb and the CIA Search for Mind Control*. New York: Henry Holt.

Kodama Muscente, Kailee. 2022. "Mandated Reporting Responsibilities as an Ethical Researcher." *Teachers College, Columbia University*. Retrieved June 17, 2023 (https://www.tc.columbia.edu/institutional-review-board/irb-blog/2022/mandated-reporting-responsibilities-as-an-ethical-researcher).

Kotsis, Sandra V., and Kevin C. Chung. 2014. "Institutional Review Boards: What's Old, What's New, What Needs to Change?" *Plastic and Reconstructive Surgery* 133(2):439–45. doi: 10.1097/01.prs.0000436846.00247.73.

Kristof, Nicholas. 2014. "Professors, We Need You!" *New York Times*, February 15.

Lahart, Justin. 2011. "Stung by 'Inside Job,' Economists Pen an Ethics Code." *Wall Street Journal*, October 11.

Lareau, Annette. 2011. *Unequal Childhoods: Class, Race, and Family Life*. 2nd ed. Berkeley: University of California Press.

Le Texier, Thibault. 2019. "Debunking the Stanford Prison Experiment." *American Psychologist* 74:823–39. doi: 10.1037/amp0000401.

Lehmiller, Justin J. 2017. *The Psychology of Human Sexuality*. 2nd ed. Hoboken, NJ: Wiley-Blackwell.

Lenza, Michael. 2004. "Controversies Surrounding Laud Humphreys' Tearoom Trade: An Unsettling Example of Politics and Power in Methodological Critiques." *International Journal of Sociology and Social Policy* 24(3/4/5):20–31. doi: 10.1108/01443330410790858.

Malcolm, Janet. 1990. *The Journalist and the Murderer*. Reprint ed. New York: Vintage.

Mead, Margaret. 1928. *Coming of Age in Samoa: A Psychological Study of Primitive Youth for Western Civilisation*. New York: W. Morrow.

Moreno, Jonathan D. 2006. *Mind Wars: Brain Research and National Defense*. 1st ed. New York: Dana Press.

National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. 1979. *The Belmont Report*. Washington, DC: U.S. Department of Health, Education, and Welfare.

Office for Human Research Protections. 2010. "Certificates of Confidentiality—Privacy Protection for Research Subjects: OHRP Guidance (2003)." *U.S. Department of Health and Human Services*. Retrieved June 18, 2023 (https://www.hhs.gov/ohrp/regulations-and-policy/guidance/certificates-of-confidentiality/index.html).

Ogden, Russel. 2008. "Harm." In *The SAGE Encyclopedia of Qualitative Research Methods*, edited by L. M. Given. Los Angeles: SAGE.

O'Toole, Kathleen. 1997. "The Stanford Prison Experiment: Still Powerful after All These Years." *Stanford University News Service*. Retrieved June 19, 2023 (https://web.archive.org/web/20111118040323/http://news.stanford.edu/pr/97/970108prisonexp.html).

Öztürk, Musa. 2018. *Sosyal Bilimlerde Etik Sorunlar* [Ethical Problems in the Social Sciences]. Ankara: Nobel Akademik Yayıncılık.

Riedel, Stefan. 2005. "Edward Jenner and the History of Smallpox and Vaccination." *Proceedings (Baylor University. Medical Center)* 18(1):21–25.

Rieder, Jonathan. 1987. *Canarsie: The Jews and Italians of Brooklyn against Liberalism*. Revised ed. Cambridge, MA: Harvard University Press.

Rothman, David J. 1987. "Ethics and Human Experimentation." *New England Journal of Medicine* 317(19):1195–99. doi: 10.1056/NEJM198711053171906.

Scarce, Rik. n.d. "Rik Scarce." *Skidmore College Sociology Department*. Retrieved June 18, 2023 (https://www.skidmore.edu/sociology/faculty/scarce.php).

Schneider, Carl E. 2015. *The Censor's Hand: The Misregulation of Human-Subject Research*. 1st ed. Cambridge, MA: MIT Press.

Singal, Jesse. 2015a. "Here's What's in Alice Goffman's Dissertation." *The Cut*. Retrieved June 21, 2023 (https://www.thecut.com/2015/07/heres-whats-in-alice-goffmans-dissertation.html).

Singal, Jesse. 2015b. "The Internet Accused Alice Goffman of Faking Details in Her Study of a Black Neighborhood. I Went to Philadelphia to Check." *The Cut*. Retrieved June 21, 2023 (https://www.thecut.com/2015/06/i-fact-checked-alice-goffman-with-her-subjects.html).

Social Psychology Network. 2015. "The Story: An Overview of the Experiment." *Stanford Prison Experiment*. Retrieved June 19, 2023 (https://www.prisonexp.org/the-story).

Solman, Paul, dir. 2011. "Oscar-Winning 'Inside Job' Director Attacks Economists' Ties to Financial Sector." *PBS NewsHour*.

U.S. Department of Health and Human Services. 2017. "Federal Policy for the Protection of Human Subjects." *Federal Register* 82(12):7149–7274.

Venkatesh, Sudhir. 2008. *Gang Leader for a Day: A Rogue Sociologist Takes to the Streets*. Reprint ed. New York: Penguin.

Vollmann, Jochen, and Rolf Winau. 1996. "Informed Consent in Human Experimentation before the Nuremberg Code." *BMJ* 313(7070):1445–47. doi: 10.1136/bmj.313.7070.1445.

Warwick, Donald P. 1982. "Types of Harm in Social Research." In *Ethical Issues in Social Science Research*, edited by T. L. Beauchamp, R. R. Faden, L. Walters, and R. J. Wallace Jr. Baltimore: Johns Hopkins University Press.

Weindling, Paul, Anna von Villiez, Aleksandra Loewenau, and Nichola Farron. 2016. "The Victims of Unethical Human Experiments and Coerced Research under National Socialism." *Endeavour* 40(1):1–6. doi: 10.1016/j.endeavour.2015.10.005.

9. ETHNOGRAPHY



If we wanted to know who handles more of the housework in married households, how would we find the answer? One way would be to interview people and simply ask them. That is exactly what Arlie Russell Hochschild did in her classic sociological study of the "second shift," her term for the work that goes on in the home after the day's work for pay is completed.

With her colleague Anne Machung, Hochschild (1989) interviewed 50 heterosexual married couples with children to learn about whether and how they shared the second shift. Many of these couples reported that they divvied up the workload equally. Some even split their residences into areas that were "her" responsibility and "his." But Hochschild wasn't satisfied with just listening to what couples had to say about all the housework they did. She chose to observe 12 of these couples in their homes as well to see for herself what housework they did.



How a couple shares the housework that needs to get done can be a major point of tension in a relationship. When Arlie Russell Hochschild interviewed couples about how they divided up household chores for her book *The Second Shift* (1989), some boasted that they split up their work equally. Yet her ethnographic observations revealed that their words didn't necessarily match their actions. *Timur Weber, via Pexels*

What Hochschild discovered was that even those couples who claimed to have an equitable division of labor fell far short of achieving it. For example, one couple who insisted that they pitched in equally to maintain their house explained that the wife was responsible for the upstairs and the husband the downstairs. After visiting this couple's home, however, Hochschild discovered that the upstairs contained all the bedrooms and bathrooms as well as the kitchen, the dining room, and the living room. The downstairs included just a storage space and the garage. This "equal" division of labor meant that the woman actually did most of the work.

The observations that Hochschild conducted at people's homes fall into a category of qualitative research called **ethnography**. Ethnography involves observing a phenomenon within its social context. Researchers—called **ethnographers** or *field researchers*—deeply immerse themselves in a particular setting, usually over an extended period of time. They collect their data by seeing what people do and say within that setting. In the social sciences, this sort of

first-hand reporting is often described as *observation*—sometimes, **ethnographic observation** (to distinguish it from the broader idea of "observation" that all scientists do when they gather data in the real world).

The setting where an ethnographer conducts observations is called a **field site**, and the act of visiting that space to do research is widely known as "going into the field." While in the field, ethnographers take extensive notes and describe their experiences in considerable detail, typically with the goal of capturing a vivid and sweeping portrait of the locations and people they observed. When they leave the field, ethnographers then focus on analyzing the rich information they gathered, often through an **interpretive approach**—making sense of what people say or do, with a focus on the social significance of those words and actions. (Confusingly, a study built on the methodological approach of ethnography is called *an* ethnography—as in, "Victor wrote an ethnography about unemployment.")



For his book *Cop in the Hood* (2009), Peter Moskos became a Baltimore police officer, successfully graduating from police academy to a job doing midnight patrols of the Eastern District's impoverished neighborhoods. His ethnographic account provided a telling portrait of the challenges faced by frontline officers, who felt bogged down by the need to interdict drug crimes and fulfill arrest quotas. This study is an example of how ethnographers sometimes choose to participate in the contexts they study—in Moskos's case, by actually doing the very profession he wanted to observe. *Kindel Media, via Pexels*

When they are in the field, ethnographers are not necessarily standing around with a notepad, mutely documenting how people go about their days. As we will describe in this chapter, they frequently participate in some fashion in the activities they observe—doing anything from volunteering for an organization, to training and working in an entirely new profession. And even if they choose to observe in the capacity of bystanders to the action, they will often ask questions or otherwise interact with the people in the setting, opening up new possibilities to gather interesting data. Their on-the-spot conversations with research participants can be instructive in themselves, but many ethnographers go a step further. Like Hochschild did in her study of married couples, they may conduct formal sit-down interviews with the people associated with their field sites—before, after, or during their observations. Combining ethnographic observation with this other qualitative method—**in-depth interviewing** (discussed in the next chapter)—gives researchers a more comprehensive understanding of the phenomenon at hand. It also allows them to use one method to evaluate or extend the findings derived from the other, in a process known formally as **triangulation**. For instance, if Hochschild had just interviewed couples rather than observing their behaviors for herself, she might have completely missed how they were saying one thing about household labor but doing something very different. Her observations helped her critique and build on what she uncovered from her interviews.

While observation and interviewing often go hand-in-hand, it's important not to confuse these two approaches. For one thing, some qualitative researchers focus mainly on in-depth interviewing (also known as *qualitative interviewing*) and do little, if any, first-hand observations of relevant social settings. (Strictly speaking, qualitative studies of this kind would be called "interview-based studies" rather than "ethnographies.") Furthermore, "ethnography" is a term that, in recent years, has widened to include

observations in digital spaces. In a **digital ethnography**, for instance, a researcher observes—as a bystander or participant—what is transpiring on social media platforms, video streams, and the like, often in real time. We'll have more to say about approaches to studying online interactions in **Chapter 15: Materials-Based Methods**. In this chapter, we'll focus on in-person ethnography—observations in real life. As you will find, this approach demands a range of widely applicable skills that relate to discerning and documenting the world around you.

Opening chapter image credit: *Marina Vitale, via Unsplash. Adapted by Bizhan Khodabandeh.*

9.1. When and How to Use Ethnography



The trailblazing anthropologist Franz Boas (1858–1942) conducted one of the first ethnographies, a report titled *The Central Eskimo* (1888) that he wrote for the Bureau of American Ethnology. For this study, Boas lived with the Inuit people on Baffin Island in Canada's Arctic Archipelago and observed their daily activities, detailing everything from their cultural beliefs to their procedures for making harpoons and igloos. *Александр Велигура, via Pexels*

Learning Objectives

1. Define ethnography and the approach ethnographers take to study a social setting.
2. Describe the participation continuum and the advantages and disadvantages of standing at any point on this spectrum.
3. Discuss the strengths and weaknesses of ethnographic observation as a research method.

Anthropologists pioneered the use of ethnography through extended in-person observations of indigenous peoples. Notably, Franz Boas (1888) lived with the Inuit on Canada's Baffin Island and then went on to study native populations in the Pacific Northwest. Bronislaw Malinowski (1922) spent years observing indigenous people living in the Trobriand archipelago of Papua New Guinea. Up to that time, anthropologists had largely studied groups of people by drawing on secondhand accounts or inspecting artifacts. The compelling findings unearthed for Western audiences by these early ethnographies, however, popularized a new style of field research that involved living in or near the communities being studied and learning directly from them.

Soon a number of sociologists adapted ethnographic methods to advance their own research projects. The University of Chicago's sociology department was incredibly influential in shaping this new approach to research in part because of its connection to Hull House, a settlement house that provided assistance to low-income immigrants (Deegan 1988). Many of sociology's first ethnographers were former social workers who became interested in such research through their personal experiences working with the poor and advocating for social reforms. They developed what eventually came to be known as the "Chicago school" of sociology, characterized by its systematic use of ethnographic observation to understand the city's neighborhoods—block by block.



Anthropologist Clifford Geertz (1926–2006) popularized the term *thick description* to describe the highly detailed and contextualized accounts that he argued ethnographers should strive to generate while in the field. Geertz himself adopted this approach for his seminal study "Deep Play: Notes on the Balinese Cockfight" (1973), which analyzed the social meanings of cockfighting in a Balinese village. While Indonesia's elites looked askance at the traditional practice and banned it, Geertz showed how central it was to understandings of status within the community. In lucid detail, he described village men wagering large amounts of money over the fights and obsessing over the care and maintenance of their animals, which became, he said, representations of their own masculinity. *Indonesian Ministry of Information, via Wikimedia Commons*

In the 1970s, the anthropologist Clifford Geertz developed an ethnographic practice called **thick description**, which would influence qualitative researchers across disciplines. If "thin" description was a surface-level observation of what people were doing, "thick" description provided needed context, allowing the researcher to understand the social and subjective meanings of particular actions (Geertz 1973). We've mentioned thick description a few times in this textbook by now, but it merits a thicker description of its own.

Thick description generates rich and complex qualitative data. Researchers using this approach document even minute details of what is unfolding before them, focusing on a micro, and sometimes microscopic, level of analysis. They pay particular attention to any patterns and processes they've observed. They capture narratives of the social interactions and relationships they've witnessed. Yet they are not intent on producing merely a facsimile, transcription, or photorealistic rendition of what they've seen or heard. (If that was the case, ethnographers could just walk around with video cameras.) They also want to interpret the underlying meanings of what they see and hear.

This is the purpose behind going into the field: to study social phenomena within their natural setting, a specific social and historical context. (For this reason, sometimes social scientists describe qualitative approaches as *naturalistic*.) Proponents of thick description (and qualitative methods more broadly) emphasize the importance of context. You cannot truly understand what you observe, they argue, without viewing it against the background in which it is situated. You will miss important nuances and misinterpret basic truths.

Being embedded in a social context and gleaning rich qualitative data from it gives ethnographers the ability to understand complex processes and relationships and explain the meanings of what they observe—that is, the social significance of specific behaviors, symbols, appearances, and interactions. This in-depth, interpretative approach is very different from the one taken in quantitative analysis, which breaks down complicated phenomena into numbers for easier processing. Through quantification, researchers can compare numerous and diverse cases and identify larger patterns across them. Yet they lose something, too, because quantification strips away context and otherwise simplifies social reality. As you can guess, these different qualitative and quantitative approaches are useful in their own ways.

In **Table 9.1** we quickly describe several celebrated ethnographies, the research questions they posed, and the types of data they collected. As we discussed in **Chapter 4: Research Questions**, ethnography—and qualitative research more generally—can answer "how" kinds of questions quite well: *how* the processes being studied occur, *how* the people being observed interact, and *how* events being witnessed unfold. In line with our discussions in previous chapters, such studies tend to adopt an inductive approach, generating plausible and nuanced theories that answer these "how" questions.

Table 9.1. Examples of Ethnographies

Author and Title	Research Question	Details about the Study
William Foote Whyte, *Street Corner Society* (1943/1993)	How is the social structure of a local "slum" organized?	Whyte was a pioneer of the use of participant observation in sociology. In this book, he spent more than three years observing a low-income Italian community in Boston's North End.
Elliot Liebow, *Tally's Corner* (1967)	How do the urban poor live?	Liebow spent two years conducting participant observation within an African American community in Washington, DC.
Michael Burawoy, *Manufacturing Consent* (1979)	How and why do workers consent to their own exploitation?	Burawoy worked as a machine operator in a Chicago factory for 10 months. He also observed other workers in the factory.
Annette Lareau, *Unequal Childhoods* (2011)	How does social class affect parenting styles?	Lareau and her research team conducted in-home observations of 12 families (poor, working-class, and middle-class), with about one month spent in each household.
Randol Contreras, *The Stickup Kids* (2012)	What changes in the economy and society drew crack dealers in the South Bronx to turn to drug robberies?	Returning to where he grew up in the South Bronx, Contreras gained access to a group of Dominican drug robbers known as the "Stickup Kids." He spent three years intensively observing the group, with less sustained observations in the years afterward.
Brooke Harrington, *Capital without Borders* (2020)	How do the ultrarich manage to avoid taxes on income, capital gains, property, and inheritance?	Over multiple years, Harrington interviewed wealth managers and trained to become one herself in order to understand the strategies the wealthy use to protect their fortunes.

One thing to note is that all of these ethnographies were published as books. Some journals (such as *Qualitative Sociology*) specialize in qualitative work, and ethnographers routinely publish portions of their work as articles in top sociology journals. However, given their use of rich detail and narratives, ethnographies often lend themselves to a monograph format. Many journals tightly control the length and structure of their articles, and ethnographers often prefer the greater flexibility they have when writing up their research as books (which can themselves be called "ethnographies"). This means, however, that the book that results from an ethnographic study may only be published many years after its fieldwork has concluded.

Now that we have a better sense of the kind of observations involved in ethnography, let's talk about the general approaches to conducting it and the strengths and weaknesses of the method.

The Participation Continuum



Richard Ocejo (2014, 2017) trained as a butcher to study how traditionally working-class jobs had become high-end professions for college was as a result of gentrification, the growth of markets for artisanal products and services, and consumers' increasing sophistication of cultural taste. *Leonardo Luz, via Pexels*

As we noted, many ethnographers choose not to be passive observers of social movements, companies, agencies, and the like. In fact, there is a storied tradition in sociology of actively contributing to the success of the organization or occupation under study—as a volunteer, a dues-paying member, or even an employee.

Because ethnographers get involved in the organizations they study so frequently, the term **participant observation** is sometimes used as a synonym for ethnography. This is a bit misleading, though, because the degree to which ethnographers "participate" in activities within the settings they observe actually varies quite a bit. You might say that there is a **participation continuum** in ethnography: complete participation lies at one end of that spectrum, and complete nonparticipation lies at the other (**Figure 9.1**).[1] A common term for the latter—not engaging in the activities or groups being studied—is *direct observation*. This phrase, however, implies a level of objectivity and intensity that, as we discuss later, is not necessarily heightened at this end of the spectrum. Alternatively, you can think of this approach as **bystander observation**, where the researcher is an eyewitness, but not a participant, in relation to the action being observed.

1. This continuum was originally described as a "participation–observation" continuum between complete participation and complete observation (Glesne and Peshkin 1992; Gold 1958; Junker 1960), but here we use the shorter term to accommodate the participation/nonparticipation gradient discussed later—and to emphasize the fact that observation of some kind is always occurring.

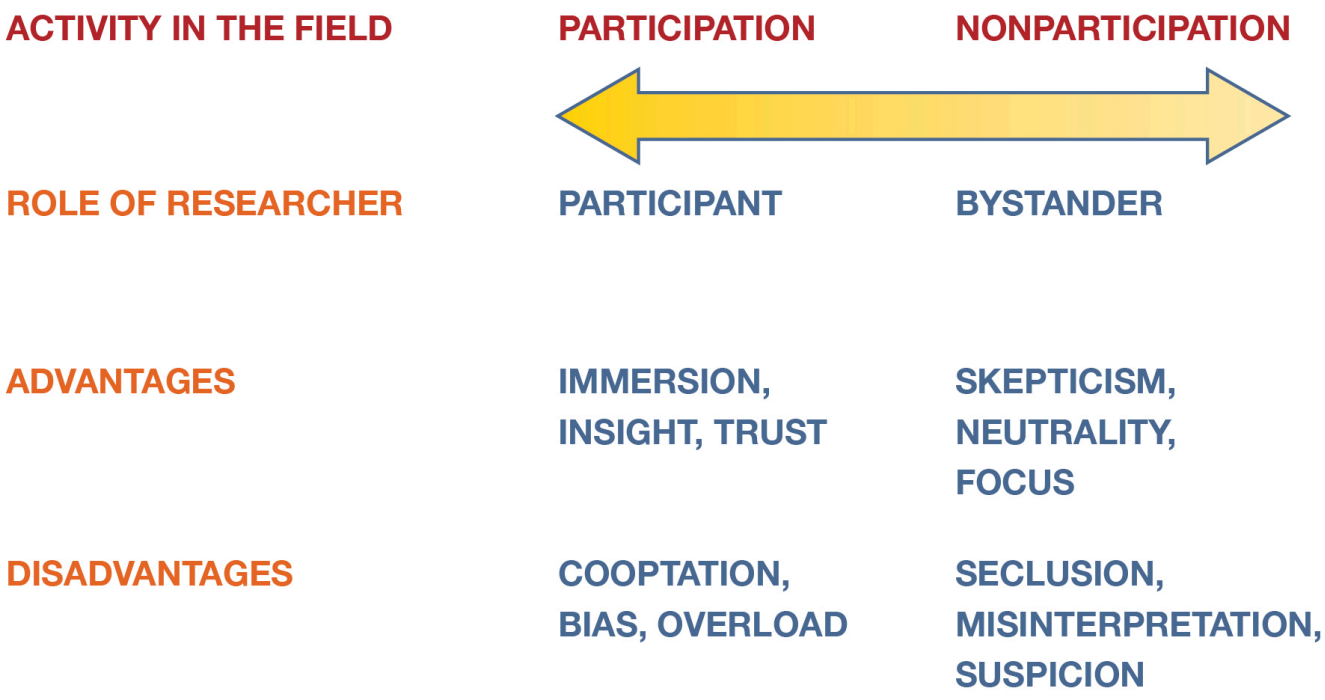


Figure 9.1. The Participation Continuum. Ethnographers can choose to be more or less involved in the activities of the group they're studying or the setting they're observing. There are pros and cons to standing at any point in this continuum between participation and nonparticipation.

In other chapters, we discussed two works that fall near either end of the participation continuum. Barrie Thorne's (1993) observations of children in classrooms, school cafeterias, hallways, and playgrounds (mentioned in **Chapter 5: Research Design**) would lie near the "bystander" end of the spectrum. Rather than taking on the role of, say, an elementary school teacher, Thorne simply observed. The diametrically opposite approach was taken by Laud Humphreys (1970) in his controversial research on the "tearoom trade," an informal but structured practice of anonymous and illicit sex taking place in men's public restrooms (described in **Chapter 8: Ethics**). As you might remember, Humphreys participated as a "watch queen," a role that actual participants in the tearoom trade also played.

One can argue that ethnographies always exist in some sort of middle ground between the poles of complete participation and nonparticipation. (And sociologists, as usual, disagree about what "participation" entails to begin with.) If you take a deep dive into the methodological debates among ethnographers, you'll find different typologies and advice for where you might locate yourself on this continuum. Some argue that true "nonparticipation" might entail something like a video content analysis (discussed in **Chapter 15: Materials-Based Methods**), where the individual observes only a video recording of an event, with no physical participation whatsoever (Johanson and Williamson 2013). For our part, we recommend that you simply recognize the many gradations of participation that exist, however you categorize them—from standing aloof from, or peripheral to, the scene being observed (with the goal of being a more objective observer), to more or less actively engaging in some activities (yet still prioritizing your role as a researcher), to becoming (or already being) a full-fledged member of the group, accepted by all. You may even decide that sometimes you'll take on one role and other times another, tweaking your degree of clinical detachment depending on what is appropriate within a specific social context.

In weighing the pros and cons of participation, it's worth considering how journalism—a trade that also does its fair

share of observation—would view such arrangements. Without question, volunteering or working for an organization you're writing about would flagrantly violate ethical codes across the more reputable news outlets within the media industry. As we noted in **Chapter 8: Ethics**, however, journalism and sociology have different professional standards regarding bias and fairness, each of which can be justified on its own terms.

Compared with journalists, ethnographers are much more willing to occupy positions closer to the participation side of the continuum. The rationale is that participant observation provides an authentic view of the life of the group being studied. Participants may trust the researcher more, and as a result, they will act more naturally, without the desire to hide certain behaviors or information from outsiders. The researcher, in turn, may gain esoteric or hidden knowledge about how the group operates or thinks, providing greater insight than an outsider could obtain. For these reasons, some ethnographers argue that participation is the *only* way to truly understand a particular social setting. Whether or not that's true is debatable, but the trade-offs of standing at a remove from those you're studying is very real. Being an outsider observer looking in on the organization may mean that you may miss or misunderstand the meanings of what people say or do around you. You may also have difficulty building rapport with the members of the group you're studying, perhaps preventing you from observing important aspects of the group's social life.

The issue with being deeply embedded in the social worlds of those you study, of course, is *ethnographic cooptation*: you may start looking with an uncritical eye upon people's beliefs, practices, and decisions. You may start agreeing with their claims or the righteousness of their cause while disregarding any contrary evidence, falling heavily into the trap of **confirmation bias**. You may forget how outsiders (the eventual readers of your published work) would look upon this group, which may be starkly different from how group members understand themselves—perhaps for good reason. These risks bring to mind a "parable-ish" joke recounted by the novelist David Foster Wallace (2009) in a commencement speech: "There are these two young fish swimming along, and they happen to meet an older fish swimming the other way, who nods at them and says, 'Morning, boys. How's the water?' And the two young fish swim on for a bit, and then eventually one of them looks over at the other and goes, 'What the hell is water?' " In a research context, you may need to step outside the water for a bit to recognize the importance of what everyone else seems to take for granted.



Some ethnographers choose to take on a more distant and dispassionate orientation toward their field site, which Fred Davis (1973) calls the "Martian" viewpoint. By that, he does not mean trying to blow up Earth in order to have an unobstructed view of Venus, but rather choosing to *not* adopt the mindset (and prejudices) of the group being studied. Much like journalists do, the "Martian" takes on a skeptical view, challenging what they are told. They try to be a fly on the wall rather than a highly visible part of the scene they are observing. *Warner Bros., via Wikipedia*

Another danger of getting in too deep (no pun intended) is that you may face heightened pressures to do something unethical or unsavory. For instance, a participant observer of criminal activities might find themselves witnessing—or even being urged to commit—something illegal. In **Chapter 8: Ethics**, we described a few high-profile cases of ethnographers who allegedly crossed one or more of these ethical lines, but we should point out that peer pressure is a common danger even when researching quite mundane organizations. Colleagues or higher-ups may push you to do things you're not entirely comfortable with at a particular workplace or protest site, for instance.

Insiderism of this sort is seen as a serious problem in the news profession, which, at its best, seeks to avoid the partisanship and credulousness that allow for the manipulation and distortion of news reporting. Similar criticisms may

be leveled against researchers who fall into this category of "going native," as this phenomenon is sometimes called. Yet ethnographers, unlike journalists, are often willing to risk such observational overembeddedness in order to reap the advantages of full participation we've just described.

A more practical reason not to participate so much in the activities of the group being studied is that bystander observers can focus just on the task of observing. Observation can be exhausting and time-consuming in itself, depending on how much detail is being gathered. Not surprisingly, when you are actually involved with the organization you are studying, you can become too absorbed with your tasks to pay sufficient attention to what is going on around you. You may also run into issues with **reactivity**, in that interactions between you and your research participants may lead to certain outcomes that differ from what would have happened if the researcher wasn't there. In this sense, nonparticipation—and even invisibility—is preferable. The British writer and aristocrat Jessica Mitford famously related how her governess would always whisper in her ear before any social occasion: "You're the least important person in the room and don't forget it" (Didion 2008). That savage putdown for an aristocrat can actually be good advice for an ethnographer, who generally does not want to be the center of attention. In fact, making yourself as inconspicuous as possible may embolden people to act more like their authentic selves in front of you.

So where should you locate yourself on the participation continuum? Some ethnographers argue that a happy medium is best: straddling the two roles of observer and participant, so that you gain greater access to the social world being studied but you refrain from "going native" (Chatman 1992). Fred Davis (1973) famously captured the equal value—and trade-offs—in these two vantage points when he described the ethnographer as being either a "Martian" or a "convert." Martians, he argued, kept some emotional and social distance from their participants to see them with fresh eyes. Converts dove deeply into the social life of their field site. One perspective assumed that dispassionate observation yielded a fuller truth; the other argued that total immersion led to true understanding. You may find yourself favoring one or the other approach in your own ethnographic observation, and you should do what fits best with your personal style. But remember the dangers of leaning too hard in one direction.



It's common sense that you don't want to lose any semblance of healthy skepticism by diving too deeply into another culture—a phenomenon known as "going native," and a constant theme in popular media. In a similar way, ethnographers who immerse themselves within the social setting they are observing may fall prey to ethnographic cooptation. They may uncritically adopt the perspective of the community being studied, which can harm the veracity and believability of their findings in the field. *Comedy Central*

The debate over participation we've just had falls into a favorite pastime of ethnographers: **reflexivity**. This is a self-reflective process that sociologists engage in to understand how their own identity, beliefs, dispositions, actions, and practices may have influenced their research. A similar concept is **positionality**, which holds that one's personal identities and perspectives shape how one interprets reality. The two terms are used somewhat interchangeably, but reflexivity is also attuned to the ways that a researcher's actions (not just their identity) may influence their findings. For instance, reactivity—the ways that your presence in a setting may alter how participants act—is something sociologists also have to consider when they are being self-reflexive about their work.

Ethnographers routinely devote significant portions of their writing to this kind of critical evaluation of themselves as researchers. In separate *reflexivity statements* or *positionality statements*, or in discussions woven into the methods sections of their papers or books, they will describe in detail the specific roles they had in the research process, what implications their involvement may have had for what they ended up observing, how their group identities may have affected their research, and so on.

This continual self-examination can help ethnographers make sure that they are getting things right—that they aren't imposing their outsider perspective upon a community they helicoptered into, for instance. It can also make them more sensitive to their treatment of their research participants and whether it is ethical and respectful. And it helps explain why many sociologists find it acceptable, and perhaps even preferable, to take on participant roles. Rather than trying to achieve some perfect state of objectivity, many sociologists would rather just acknowledge their biases and discuss

what consequences they may have had for their analysis. (We'll have more to say about reflexivity when we talk later about how an ethnographer's personal identity affects their research.)

Regardless of whether you choose to observe as a participant or a bystander, remember that as a qualitative researcher, you are part of the measurement instrument. So much depends on your mettle as an individual existing in a particular social moment. In the field, you must rely on your observational skills, the trust you build with participants, and your ability to extract the correct information. Your highly subjective and personal experiences in the social context you are studying will allow you to interpret what is going on in a rich and full fashion. At the same time, if your goal is scientific—rendering an accurate portrayal of the phenomenon you're studying—you need to maintain a fair-minded perspective. Keep your biases in check so that you don't—consciously or unconsciously—twist the details you see to fit whatever narrative you prefer. Confirmation bias is very real.

The Strengths and Weaknesses of Ethnography

Ethnography allows researchers to gain firsthand experience and knowledge about the people and processes they study. Perhaps no other method offers an experience of everyday life that is this close-up and detailed. As we have noted, ethnography is a superior method for understanding the role of social context in shaping people's lives and experiences. It excels at illuminating causal mechanisms as well—the specific pathways by which one variable influences another—given that thick description can fully capture complex relationships and interactions. In fact, as an ethnographer, you may be able to see these processes unfold before your very eyes. Few kinds of data have greater face validity than what you can gather here: a clear and cogent story, directly observed.

Ethnography is also an excellent method to use for exploratory research. It is likely to generate results that surprise you, teaching you something about the phenomenon you're studying that you didn't know before you went into the field. With surveys and interviews, you have to know what questions to ask other people about a phenomenon of interest; with ethnographic observation, you can see it for yourself, which can draw your attention to things that the people regularly in those settings may not think about themselves. Furthermore, the freedom you have as a researcher to keep exploring the same sites or interacting with the same sorts of people can allow you to adapt to new information acquired in the field, test out your latest hunches, and seek out theoretically interesting cases (in the **abductive approach** we described in **Chapter 4: Research Questions**). As an ethnographer, you should take full advantage of the method's nonjudgmental flexibility and openness to new possibilities. "I had a mentor who taught me that if you end up with the story you started with, then you weren't listening along the way," says the documentary filmmaker Matthew Heineman (McClennen 2022). "It's good advice for life and good advice for filmmaking." It's also good advice for ethnography.

The unmediated access that ethnographers have to the phenomena they're studying also allows them to avoid some of the measurement problems that bedevil other methods. For instance, in-depth interviews and surveys ultimately depend on the hopeful notion that respondents will accurately relay their experiences and reasons for doing things. But what if their recall is faulty? What if they don't know why they really do what they do? Or what if they feel uncomfortable about telling you what they truly think? Speaking again to this idea of being pleasantly and productively surprised, ethnographic observation can sometimes tell us things that our research participants don't even know. And because ethnography typically occurs over an extended period of time, the researcher may discern certain social facts over the course of their study that were at first invisible or hidden from them.

Broadly speaking, ethnography involves two challenges. First, ethnographers need to be fully engaged with the social context they're studying to generate useful data and theories. Ethnographic observation is a very labor-intensive form of research that requires researchers to think creatively and improvise constantly while in the field. It often takes a great deal of time to collect sufficient data on a few cases—even just a single site of study. Documenting ethnographic observations can also suck up a great deal of time and energy. In interview and survey research, respondents typically enter their responses themselves or have their spoken words recorded. Ethnographers generally have only themselves

to rely on for documenting what they observe. This challenge becomes immediately apparent upon entering the field. As we discuss later, it may not be possible to take field notes while you are observing. Afterward, you won't recall everything exactly as you saw it. Nor will you necessarily know which details to document out of the countless number that you could write down. The most revealing bits of ethnographic data for your study may very well turn out to be ones you thought were inconsequential when you first recorded them. This is the chaos and uncertainty that ethnographers regularly deal with—and that good ethnographers learn to take in stride.



One or more interactive elements has been excluded from this version of the text. You can view them online here: https://viva.pressbooks.pub/sociology-research-methods/?p=300#oembed-1

Video 9.1. *Sidewalk*, the Documentary. A decade and a half after sociologist Mitchell Duneier wrote his celebrated ethnography *Sidewalk* (1999), he returned to Sixth Avenue in Greenwich Village to update the stories of the unhoused street vendors he profiled in his book. The result was a new documentary collaboration with filmmaker Barry Alexander Brown and photographer Ovie Carter.

Ethnography can also be emotionally taxing. As with in-depth interviewing, ensuring your observations yield useful data often requires establishing a rapport with the people you are studying; otherwise, participants may hide things or refuse to be part of the study in the first place. And while interviews and ethnography both require building such relationships, you might say that in-depth interviewing is like casual dating, while ethnography is like a long-term relationship. The latter generally requires a narrower focus and a heavier commitment—one often sustained over a much longer period than the hour or two it might take you to conduct a typical interview. As a result, the relationships you develop as an ethnographer tend to be more intimate. A number of ethnographers have described their close bonds—and, sometimes, complicated relationships—with research participants (Duneier 1999; Lareau 2011; MacLeod 2008). Gaining the trust and respect of your participants not only makes it more likely that your observation will yield rich data, but it can also be personally rewarding. It's hard not to spend so much time with other people and not feel some connection to them and some enjoyment of their company. As in any relationship, however, you will experience emotional highs and lows, as you encounter the best and worst of the people and organizations you are studying. For a ground-level perspective of what this process of emotional immersion is like, watch **Video 9.1**, a companion documentary for Mitchell Duneier's book *Sidewalk* (1999).

The second overarching challenge facing ethnographers is one we've described before (including in **Chapter 6: Sampling**) as a pitfall for all qualitative researchers: generalizability. Compared with other methods, it is a much heavier lift for ethnographers to convince other researchers that the social patterns they observed within their chosen setting or settings somehow apply more broadly. This is inevitable given the first drawback we discussed, the labor-intensive nature of ethnography. The approach is so detailed-oriented that it is by necessity somewhat narrow. Ethnographers are forced to focus on one or a handful of field sites, restricting themselves to a quite limited view of a larger phenomenon. They simply cannot gather data about as many individuals or locales as, say, a survey researcher can. Of course, this limitation of ethnography is also its strength, because ethnographers sacrifice breadth in exchange for depth. They retain the stories, context, and meanings that quantitative analyses by necessity strip away (Wang 2016). From the fine-grained and richly contextualized data their observations generate, they can piece together highly nuanced and sophisticated understandings of causal mechanisms, subjective viewpoints, and the underlying social meanings of what people do and say.

Key Takeaways

1. Originally developed by anthropologists, ethnography involves immersing oneself in a field site over an extended period of time and observing the social meanings of what people do and say within that setting. Unlike journalists, ethnographers frequently participate in the activities or organizations they are studying. This participant observation can be contrasted with direct observation, or bystander observation, where the ethnographer is observing as a bystander to any activities.
2. There are advantages and disadvantages associated with locating yourself on any point of the participation continuum. Participant observers can become so immersed in the work and culture of the group they're studying that they have trouble critiquing its practices or perspectives in a fair-minded fashion. Yet bystander observers face a greater risk of being shut out by the group or misinterpreting the actions they see or words they hear.
3. Ethnography is especially good at yielding detailed data, placing people's views and actions in their social context, and uncovering surprising findings. However, making observations and documenting them can be physically, logistically, and emotionally taxing. As a result, ethnographers can typically study only one or a few sites and produce results that may not be generalizable to the target population.

Exercises

1. Where do you think is the best place to reside on the participation continuum? What are the advantages and disadvantages of being at either end?
2. Find an article reporting results from an ethnographic study. (You can do this by using any of the databases described in **Chapter 5: Research Design**.) How do the authors describe the strengths and weaknesses of their study? Do they touch on the pros and cons we've just discussed, or do they flag other issues?

9.2. Going into the Field

Learning Objectives

1. Discuss grounded theory and the extended case method and how ethnographers utilize these approaches when entering the field.
2. Describe the factors you should consider when choosing and gaining access to a field site.
3. Explain the role of reflexivity when conducting observations in the field.

As we mentioned in **Chapter 4: Research Questions**, qualitative research typically involves an **inductive approach**—put simply, connecting the dots between aspects of different observations, and seeing what theories emerge as a result. Sometimes ethnographers (and other qualitative researchers) conduct a kind of "pure" inductive analysis called **grounded theory** (Charmaz 2006; Glaser and Strauss 2017).[1] In this approach, researchers begin with an open-ended and open-minded desire to understand a social situation or setting, letting the data guide them rather than guiding the data with a set of hypotheses in mind. The goal when employing a grounded theory approach is to generate theory. Its name not only implies that discoveries are made from the "ground up," but also that any theories that emerge are actually grounded in a researcher's empirical observations and a group's tangible experiences, rather than being merely theorized into existence. In fact, to ensure that the theory is based solely on observed evidence, the grounded theory approach asks that researchers put aside their preexisting theoretical blinders while they are in the field and, afterward, letting the data alone dictate the theory that emerges.

As exciting as it might sound to generate theory from the ground up, the experience can also be quite intimidating and anxiety-producing, as the open nature of the process can sometimes feel a little out of control. Without hypotheses to guide their analysis, researchers engaged in grounded theory work may experience feelings of frustration and angst. For that reason, many ethnographers take hybrid approaches that stand somewhere between grounded theory and hypothesis-driven social science. One popular method that creatively uses existing theory as a departure point is the **extended case method**, pioneered by the ethnographer Michael Burawoy (1998, 2009). Generally speaking, this strategy involves finding a theory that somehow fails to explain what is occurring in a particular social context. The ethnographer then embarks on a case study, conducting observations within that site to understand why the theory falls short—and how it can be "rebuilt" to accommodate this deviant case. Burawoy pursued this approach in his ethnography *Manufacturing Consent* (1979), a groundbreaking study of factory workers. Marxist theory predicted that workers would organize to resist their oppression, but Burawoy described how workers' energies were instead diverted to a "game" of meeting production targets, which blunted any collective resistance to management.

The general approach you decide to take to qualitative research—essentially, how inductive or abductive your inquiry will be—will, in turn, shape two practical decisions you need to make before entering the field: where you will observe

1. You can learn more about grounded theory at the Grounded Theory Institute, a site run by one of the pioneers of the approach, Barney Glaser.

(i.e., your field site) and what role you will take while in the field. Let's talk about these considerations and what other factors will figure into your plans for fieldwork.

Choosing a Field Site

The location you observe might be determined by your research question, but because ethnography is often exploratory and inductive, it can be acceptable to jump into your observations even before you have settled on a focused question. You may enter the field with just a vague notion of the topics you want to study, homing in on a specific set of research questions after you have collected data. Even if you begin with a pointed question in mind, you should remain open to the possibility that your focus may shift as you see what happens in your field site—remember, one of the strengths of ethnography is its ability to surprise you.

Regardless of how focused you are at the outset of your ethnographic study, you should consider the following when choosing a site to observe:

1. What do you hope to accomplish with your research?
2. What is your topical or substantive interest, and where are you likely to observe behavior that has something to do with that topic?
3. Will your observations be limited to a single location, or will you visit multiple locations?
4. How likely is it that you'll actually have access to the locations that are of interest to you?
5. How much time do you have to conduct your observations?

Questions 2 and 3 should make you think of the discussion in **Chapter 6: Sampling** about purposive sampling. You should choose a field site that will give you a useful window into the phenomenon you're studying. Maybe you want to choose two or more sites that differ in key ways, so that you can compare how that variation influences what you see in each setting. The site or sites you pick do not necessarily have to be "average" cases of your population of interest—say, an "ordinary" neighborhood in a particular city or a "typical" company in a particular industry. Choosing such a site might make it easier to justify the importance of your study, but as we've said before, with such a small sample—at most, a few sites—it's foolish to focus so much on finding a perfectly "representative" case. What's more, an "average" site may be utterly boring, lacking the kinds of variation that qualitative researchers often find theoretically profitable to explore.

Depending on your research question, you might instead want to study **extreme cases** of the kind we discussed earlier in the book: a field site that vividly and intensely displays a particular concept of interest. For example, Katherine K. Chen (2009, 2012) chose to study Burning Man, a unique organization that puts on an annual arts festival in the Nevada desert using a deeply democratic form of decision-making and decentralized management. As Chen points out (2015:38), organizations with extreme characteristics can "render visible the invisible," highlighting for us practices and beliefs that we take for granted, such as "the concentration of power in our everyday organizations, particularly hierarchical ones, and our acceptance of this as normal or even desirable." By studying an organization without such a concentration of power, Chen put a spotlight on the taken-for-granted work that goes into making an organization run smoothly without stark hierarchies and lots of rules—work that would be much harder to see in an "average" organization.

Questions 4 and 5 get at some of the practical issues that will surely influence which field site you choose to study—or even whether or not you decide to conduct an ethnography. First, the ideal locations to observe may not be available to you. If you're not observing a public space like a street or park, you will need permission to even step foot onto the property as a researcher. Schools, government agencies, and corporations may refuse you outright or require you to go through an elaborate and uncertain approval process. If you don't have a person in the organization to vouch for you, you should expect doors to be slammed in your face many times before a single group agrees to consider your request. Note, too, that if an organization extends permission at first but later revokes it, your previous data collection may be compromised, depending on what arrangements you worked out with the group and with your **institutional review**

board (IRB). It may be good to get the organization's permission in writing (see **Figure 9.2**)—which some presses will want to see before they publish your ethnography, anyways.

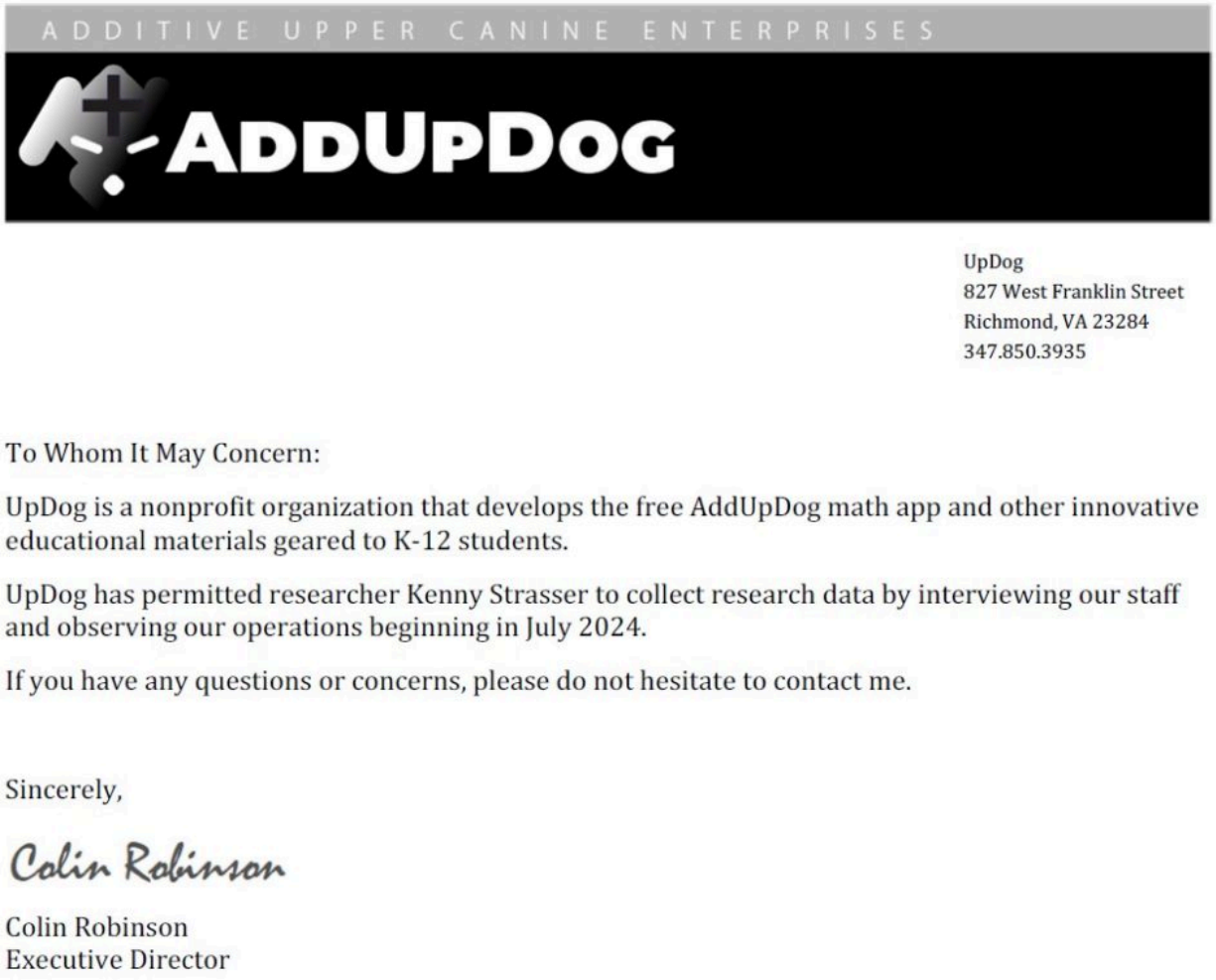


Figure 9.2. Organizational Permission Letter. If you are conducting an ethnographic study focused on a particular organization, you might want to ask the organization for a letter or other documentation stating that you have their permission to study the group. This is a sample letter from a fictional organization (based on an actual letter) providing such permission. *Elijah Kai Chen*

If you have a personal connection to any potential field sites, you shouldn't be shy about tapping those existing contacts or commitments. Perhaps you are already a member of an organization that you'd like to study. Maybe you know someone who knows someone else who might be able to help you access a site. Perhaps you just have a friend you could stay with in that city, enabling you to conduct your observations away from home. While you do need to consider how fair-minded you can be, it can be both fun and personally meaningful to study a group or cause you care deeply about. As Joyce Rothschild pointed out in **Chapter 4: Research Questions**, that insider knowledge is often your "comparative advantage" as an ethnographer, allowing you to write something that another sociologist couldn't.

Another key consideration when choosing a field site is your available resources of time and money. As we've described, ethnographers take months or even years to collect data at their field sites. Ideally, they are available to visit their site frequently, even daily, so that they get used to the setting—and so that people in that setting get used to them. (In short, ethnography is not a great method to use if you're looking to do a quick and painless study.) You need to be honest with yourself about how much time you can devote to observing your field site. You also need to think about where you live and whether regular travel to your site is an option for you. Some ethnographers actually relocate so that they can live with or near their population of interest—take Matthew Desmond, who moved into a trailer park and then a rooming house in Milwaukee so that he could study low-income housing issues for his Pulitzer Prize–winning book *Evicted* (2016). One of the exciting things about ethnography is being able to go to far-flung places and immerse yourself in a previously unfamiliar social world. (In short, ethnography *is* a great method to use if you are an adventurous person!) Of course, you can still find creative ways to do ethnographic research even if you don't have the flexibility or funds to pick up and move to the ideal field site for what you want to study.



In their book *Paying for the Party*, Elizabeth Armstrong and Laura Hamilton (2015) studied sorority members at a Midwestern public university. While they relied primarily on in-depth interviews, they also included observations (largely done by Hamilton, then a graduate student, who could navigate campus social spaces more easily). The book provides an example of how a qualitative study (even one with an ethnographic component) can be collaborative. *Henderson State University (Arkansas), via Flickr*

Collaborating on an ethnographic project might help you address these logistical issues. If you take on one or more

research partners, you might be able to cover more ground and therefore collect more data than you could on your own. A research partner could study an entirely different field site (even one in a different part of the world), significantly extending the scope of your study and leveraging a potentially interesting comparison. Having collaborators is especially helpful in ethnography because of the difficult logistical, interpersonnel, and ethical issues you are likely to encounter. You might find it useful to talk through problems or just commiserate with a collaborator who shares at least some of your experiences in the field. Of course, collaborative research also comes with its own set of challenges: possible personality conflicts among researchers, competing commitments in terms of time and contributions to the project (a particularly salient issue in ethnography, given the level of commitment required), and the sociologist's version of "creative differences"—having clashing methodological or theoretical perspectives that make it difficult to agree on how best to collect data or interpret it (Shaffir, Marshall, and Haas 1980). You should also be aware that building rapport with the community you're studying might be harder with two or more outsiders involved rather than just one.

Presenting Your Identity as a Researcher

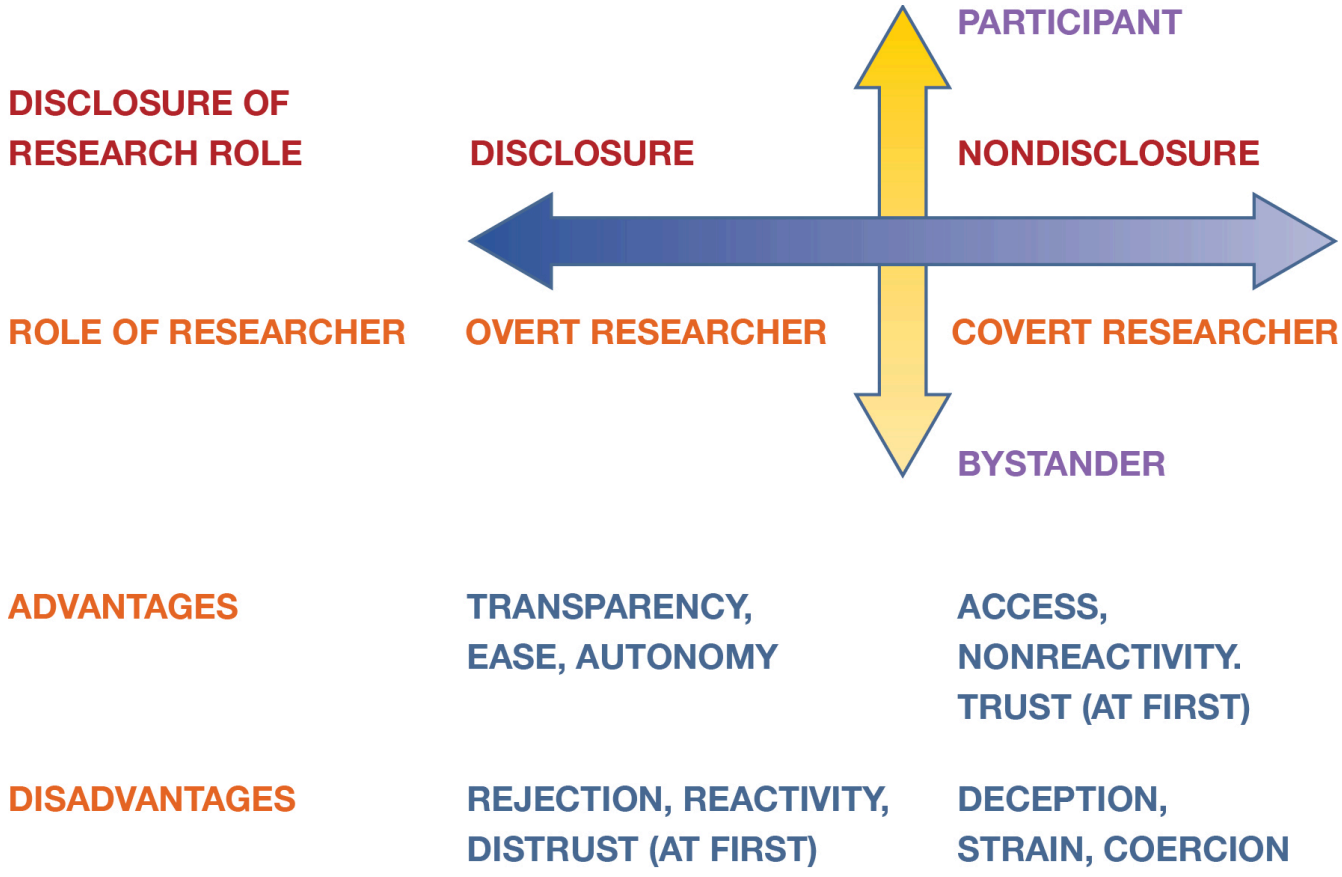


Figure 9.3. The Disclosure Continuum. In addition to deciding on how much to participate in the activities they're observing, ethnographers need to figure out the level of disclosure they want to provide to participants about what they are doing in the field. Although IRBs are resistant to covert research, some ethnographers have produced excellent studies that might not have been otherwise possible if their identities as researchers had (at least initially) been known to those they were studying. Even if your research is overt, you can choose to adopt varying degrees of transparency and explicitness about your researcher role while in the field. And as with the participation continuum, standing on any point in this disclosure continuum presents advantages and disadvantages.

Earlier in the chapter, we discussed the participation continuum and how ethnographers make conscious choices to fall somewhere in that spectrum between being a complete participant and complete bystander. Beyond that general orientation toward the people you are studying, you also have to decide the extent to which you will divulge to them your identity as a researcher. This can also be seen as a spectrum—a **disclosure continuum** (**Figure 9.3**). On one end, you hide your identity entirely (**covert research**). On the other, you make it known to everyone you observe that you are a researcher (*overt research*).

In general, not disclosing your identity as a researcher is acceptable when you are observing people in the street or in other public spaces where they have no expectations of privacy. Some ethnographers, however, may decide to be covert while observing in private settings, too. The degree to which they actively hide their identities can differ, with an entirely covert role usually reserved for studying a highly exclusive organization. In this extreme case, they may pretend to be nothing more than participants, and they may deny they are researchers if they are asked.

Entirely hiding your identity from your research participants may seem appealing because being undercover avoids some of the practical issues we have described throughout this chapter. For instance, some organizations don't like to be studied, so being open about your observer role means their leaders may refuse your requests to enter those settings. Furthermore, if specific participants are aware that you are a researcher, you may experience some trouble establishing rapport with them, especially when you are dealing with a very closed or tight-knit group of people. You can address this lack of trust by being more of a participant rather than a bystander, or by having an insider at the site who can vouch for you. Inevitably, however, people who know they are being observed will respond in some way to that observation, leading to reactivity that may skew the results of your observation. Depending on the situation, participants may be uncomfortable or even hostile in your presence. Likewise, they may behave differently than they otherwise would have out of desire to maintain a socially acceptable persona, conceal specific unflattering facts about themselves or their group, or even seek your approval as a researcher. If you choose to conduct covert research, all these issues may be mitigated.

As you probably noted, many of these risks are similar to the ones we discussed regarding the degree of participation you take up in the field. Yet it is worth distinguishing the disclosure continuum from the participation continuum in your head because they are conceptually distinct, and different combinations of participation and disclosure are possible. For instance, you can be a bystander observer who doesn't disclose your identity as a researcher, or you can be a participant observer who is utterly engaged in the organization being studied—though everyone there knows about your research.

Furthermore, falling on the covert side of the disclosure continuum presents a number of unique ethical and practical issues. As we discussed in **Chapter 8: Ethics**, it is critical for researchers to obtain informed consent from their research participants, which includes not deceiving them about the nature of the research—or about the fact that research is taking place, period. The IRB at your school or organization that vets your proposed project may let you move forward with such a study, but expect a lot of pushback. IRBs generally do not allow researchers to deceive respondents unless the rationale is highly compelling and outweighs the harm caused by not obtaining fully informed consent from the outset. (Again, more discretion is allowed when the field site is a public space where people may not have an expectation of privacy.)



For his book *Dancing at Armageddon* (2002), Richard G. Mitchell, Jr., spent a dozen years covertly observing survivalists, doing everything from attending their conferences to visiting their training camps. His deep cover allowed him to provide an extraordinary glimpse into a secretive subculture and its scathing critiques of modern industrial society. However, he later regretted participating in the swapping of racist and homophobic stories that occurred as survivalists bonded with each other (Mitchell 1991). *Jordan Bergendahl, via Pexels*

Another obvious problem with being a covert researcher is that it introduces a great deal of stress into your field research. You will constantly be worried about being exposed as a researcher and facing the wrath of deceived participants, which further adds to the psychological burden of continual observing and/or participating. Even if this strain is manageable at first, your initial decision to conduct covert research may complicate matters later as your project unfolds. If you choose to conceal your identity, for example, how long will you do so? After all, eventually your work will be published, and then the people you deceived will likely figure out your identity. Should you disclose this fact sometime before then? How might participants respond when they discover you've been studying them? If they find out before you've had a chance to tell them, they may be even more upset.

A final issue you might encounter in hiding your identity is perfectly summed up in a *Far Side* cartoon. It depicts a "researcher" dressed up like a gorilla, hanging out with a few other gorillas. In the cartoon, one of the real gorillas is holding out a few beetle grubs to the researcher, and the caption reads, "So you're a *real* gorilla, are you? Well I guess you wouldn't mind munchin' down a few beetle grubs, would you? In fact, we wanna see you chug 'em!" As a covert researcher, how will you respond if asked to engage in activities you find unsettling or unsafe? For instance, while conducting covert research among right-wing survivalists, Richard Mitchell (1991) was asked to participate in the

swapping of violently racist and homophobic stories. He later expressed regret over his involvement in those activities. The degree of coercion you are subject to may be greater if you are a covert researcher, given that you do not have the excuse of "I'm a researcher."

As you can see, the decision to study a site covertly is not as convenient as it might seem at first glance. That said, you should also note that there is some variation in the degree to which researchers flag their identity as researchers even when they are not seeking to be entirely covert. This is because there is some leeway for not *volunteering* the fact you are observing someone. Let's say an organization approves your request to study them. Ideally, the leaders let everyone in the organization know that you are a researcher. At the private meetings of the group that you attend, you can take the extra step to announce your role so that no one is blindsided. Undoubtedly, however, there is going to be more or less awareness of your researcher identity across members of the group. You will have to decide whether you will be more active or passive in making people aware—another reason to think about this process of disclosure as a continuum, rather than an either/or state. For instance, if you happen to observe a large gathering of that organization—say, a protest march they organize—it will be difficult or awkward to convey your role to everyone in attendance, and there wouldn't really be an expectation of privacy given the public nature of such an event, either. As is often the case in the field, you as a researcher will need to navigate these more nuanced situations that can come up, sometimes unexpectedly, in the course of your observations. You should do so with due consideration to both ethical principles and practicality—while still following the bright lines that your IRB has set for you in terms of getting explicit approval from those you observe.

Here's a final point we want to make about being covert: it's overrated. With persistence, you'll likely find some group willing to let you openly observe them. And although your known presence as a researcher undoubtedly changes how people act, that effect will lessen with time—another reason that ethnographers spend so long in the field. Specifically, your continual presence and constant vigilance as an ethnographer will, with time, bore even the most circumspect research participant to the point they return to acting like their usual self. Putting up a facade takes some effort, and especially if you are a participant observer, eventually you will become just another part of the scenery to them. (After all, even documentary film crews wielding cameras and boom microphones somehow get their subjects to lower their guard after, say, a few weeks of shadowing them.) Our advice for new ethnographers is patience. Stay calm and keep observing.

Considering Your Identity in the Field

In addition to managing their professional identities in the field, ethnographers also need to manage their personal identities—that is, seriously considering how they will interact with the people they encounter in their field site, and what implications the identities they possess or present will have for their research.

First, as part of the self-critical process of reflexivity mentioned earlier, ethnographers should contemplate how their social location—their identity relative to the people they're studying—might affect their ability to conduct research at a particular field site. When assessing your positionality, it's helpful to distinguish between the ascribed and achieved aspects of your identity. A*scribed* aspects are involuntary, such as your age and race, and you are stuck with them—even though they may make it easier or harder for you to conduct observations in certain settings. For instance, if you were studying racial discrimination in a particular workforce, you would want to think about how your own racial or ethnic identity might enhance, inhibit, or otherwise affect what you see or how you interpret it. Among other things, individuals from a different background than your own might be wary about what they say around you, and you might not understand the nuances of how people of a different group process the treatment they receive. If you were observing children's birthday parties, you would need to consider how being an adult (assuming you are not some sociological wunderkind) will affect how the partygoers or their parents react to you. Your very presence as a grown-up might change how the children play around you, for example.

Ethnographers do have some control over the *achieved* aspects of their identities, which would include the clothes they wear and the mannerisms they display. For instance, they may be able to hide the fact that they come from a middle-class background when they are studying poor or wealthy households (or vice versa). As with the hiding of one's identity as a researcher, the ethical line here can become a bit blurry, but ethnographers do tend to conceal or mute certain characteristics to fit in when they are in the field. They recognize that sticking out in a particular social context can make it less likely that the community at hand will welcome them into that space or trust them with insider knowledge. They also don't want people to act differently because they know an outsider is watching them. It's worth emphasizing that whether a particular identity is a benefit or hindrance really depends on the context. For example, in some cases, sharing the fact that you are a student might enhance your ability to establish rapport with participants; in other field sites, it might stifle conversation and connection.

As an ethnographer, you will need to decide on the persona you present to other people while observing them. Not all parts of your identity are equally important, and you may find it quite appropriate to play up some aspects and not others. (For instance, student researchers differ on the extent to which they share this aspect of their identity with other people at their field site.) In general, however, we would counsel you to be authentic. Doing so will keep you from misrepresenting yourself in an unethical fashion, and you'll be surprised how little people ultimately care about who you are—so long as you are respectful and spend sufficient time with them to gain their trust and acceptance.



The legendary American cultural anthropologist Margaret Mead (1901–1978) became a public figure after publishing ethnographies that described very liberal attitudes toward sex within traditional South Pacific and Southeast Asian cultures. (Her graduate advisor at Columbia University was Franz Boas, mentioned earlier as one of the pioneers of the ethnographic method.) Mead's work later informed debates about feminism and sexuality during the sexual revolution of the 1960s, but they also drew a backlash, with critics alleging that she had adopted an overly credulous and ideological position as an outsider studying cultures that were not her own. Later assessments discredited many of these criticisms (Shankman 2009), but a 2016 survey of anthropologists nonetheless found that two-thirds believed Mead had "romanticized" the sexual freedom of the cultures she had studied (Horowitz, Yaworsky, and Kickham 2019:678). *Smithsonian Institution, via Flickr*

As a self-reflexive ethnographer, you should think hard about how aspects of your identity may influence the ways that people treat you in the field, and what access you are able to obtain to the interactions and individuals you wish to study. Raising these questions does not necessarily mean that you should not study those who are different from you, however. Countless insightful ethnographies have been conducted by researchers who were not part of the populations they chose to research. Sometimes, the fact that you *don't* fit in can make it easier to navigate a particular social context. For instance, being an outsider may mean you are not perceived as a threat to either side in an organization's ongoing power struggle. As we alluded to earlier with our fish story, keeping a distance may also make it easier to see certain aspects of a culture that insiders take for granted.

That said, ethnographers today are particularly outspoken about the need for more ethnographies to be done by people who actually hail from the communities being studied. It is a sensitive subject given that ethnography was pioneered by white researchers who chose to observe "exotic" indigenous populations—and, at times, allegedly misinterpreted them. As many sociologists see it, studying a community that isn't your own adds to your obligation to get things right. You need to cultivate cultural sensitivity and avoid any glib presumptions that you truly understand an experience or viewpoint that isn't your own.

Having an insider at your site who you can turn to in order to interpret your observations can be helpful in reducing this sort of cultural disconnect. Such insiders are called **key informants**. Perhaps you already have a connection to a person within the organization being studied, or perhaps you immediately hit it off with a particular individual you happen to encounter in the field. This informant can provide a framework for your observations, help you "translate" what you observe, give you important insight into a group's culture, and assist you in recruiting additional participants to observe or interview. If possible, having more than one key informant at a site is ideal, as one informant's perspective may vary from another's—and sometimes an informant may have an incentive to portray their organization in a particular light (particularly if that person represents one side in an internal power struggle). One of the most notable key informants in sociology was the gang leader "Doc" from William Foote Whyte's classic ethnography *Street Corner Society* (1943/1993). Critics of the book wondered, however, if Doc provided Whyte with a lopsided picture of the dynamics on the North End's streets. Subsequent efforts to pin down the actual story have challenged whether Whyte faithfully conveyed Doc's take on the neighborhood to begin with (Boelen 1992).

Although we have talked about the various ways your personal identity affects your research throughout this section,

we want to conclude by cautioning you about strategizing too much about your presentation of self in the field. Multiple approaches can work (or not work). Ethnography is more of a marathon than a sprint, and you may make inroads into a social circle on one day, only to see the rapport you built disintegrate on another day. One basic piece of advice is to follow your gut—and your gut sense of right and wrong—when choosing how to interact with other people in the field. Participant observation is a demanding endeavor, and there are sometimes no easy or straightforward answers to the strategic and ethical challenges you'll encounter. The main thing is to treat others ethically while advancing your research.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://viva.pressbooks.pub/sociology-research-methods/?p=307#oembed-1

Video 9.2. *Street Corner Society*, in Retrospect. This is a 40-minute interview with William Foote Whyte and several of his research participants nearly 40 years after the publication of his groundbreaking ethnography *Street Corner Society* ([1943] 1993).

Ethnography as Empathetic Understanding: A Q&A with Amy L. Best



Amy L. Best is a professor of sociology at George Mason University who studies youth and gender identity formation, social inequalities, and school and consumer culture. She has written three ethnographies that immerse readers into the social worlds of teenagers and young adults. For her book *Prom Night: Youth, Schools, and Popular Culture* (2000), which received a Critics' Choice Award from the American Educational Studies Association, Best hung out at high school proms and described how kids worked through their understandings of authority, class, and gender through that storied annual ritual. For *Fast Cars, Cool Rides: The Accelerating World of Youth and Their Cars* (2005), she frequented car shows, car dealerships, and car washes around San Jose, went car cruising alongside throngs of young people, chatted up kids in high school parking lots, and sat in on their auto shop classes, capturing in rich detail how cars figure into their everyday lives. And for *Fast-Food Kids: French Fries, Lunch Lines, and Social Ties* (2017), which received the DC Sociological Society's Morris Rosenberg Award, Best spent months observing public school cafeterias and fast-food restaurants, seeing how high school students across socioeconomic lines used these spaces for sustenance—both nutritional and social.

In addition to her research, Best regularly engages in public debates and policy discussions relating to young people. She has spent the past decade evaluating school and community programs to improve children's consumption of fruits and vegetables and expand their access to healthy and nutritious foods. At George Mason University, she is currently codirector of the Youth Research Council and a faculty affiliate of the Center for Social Science Research. Along with Patricia White, Roberta Spalter-Roth, and Kelly Joyce, she was a principal investigator on a National Science Foundation–funded study examining how social science research informs the policy process. An expert on program evaluation based on observation, Best has also contributed to scholarly conversations about qualitative and feminist approaches to social research, editing the volume *Representing Youth: Methodological Issues in Critical Youth Studies* (2007) and coediting the New York University Press book series *Critical Perspective on Youth*.

What drew you to study sociology and become an ethnographer?

Like a lot of sociology majors, my love of sociology began in an undergraduate sociology classroom. Sociology's concepts (there are so many!) helped me to understand the world in a way I had never before. Sociology provided me with a lot of answers. It also sparked a curiosity about how things work and fit together that continues to propel my research and teaching.

When I was in college, I knew that after I graduated I wanted to focus professionally on creating positive and lasting social change, but the path to get there was less clear. I thought I might become a social worker or a lawyer. With the encouragement of several professors, though, I decided to pursue a PhD in sociology instead. By then I had engaged in two ethnographic projects. I loved the process of fitting pieces of a puzzle together, trying to make it whole. I loved the idea of understanding a larger world by examining people's daily experience of it. And while I appreciate the value of statistical research, it was the deeper kind of investigation ethnographers do that set me on a path to become a professional sociologist.

You've shown a remarkable ability to immerse yourself in social worlds that many researchers might feel self-conscious or awkward about observing—from high school proms, to car-cruising sites, to school cafeterias. How do you go about entering these spaces and gaining the sort of trust, respect, and inconspicuousness you need to do your work well? Is it harder to study youth as researchers grow older, or are there advantages to that social distance?

I am laughing a little about the first part of the question because even though I have been able to immerse myself in these worlds, I still feel self-conscious and sometimes even a little out of place. I have always tried to use those feelings to better understand the setting and the people in it. This type of ongoing reflection is a critical part of an ethnographer's toolkit. Checking our own assumptions and feelings is important if we're going to understand what's going on in a place.

I will also say as I have grown older, connecting with teenagers has become less automatic. It requires more effort to meet them where they are. When I first began studying youth and youth culture, I was in my 20s and still a student myself. It was easier for them to see me as being like them. I'm too old for that now. That said, I still find it easy to talk to teenagers. I find their energy enlivening, appreciate their penchant for physical humor, and admire the deep level of support they give to their friends. Teenagers are clever and creative—and so very funny. As a sociologist, I am always interested in decoding the expressive culture that defines youth worlds and setting it against a larger backdrop of both the past and the present.

Generally, people are usually pretty good at sensing when someone likes them. I find teenagers to be really tuned into whether someone is being respectful or patronizing. As a researcher, it's really important to be keyed into that. It's also important to be able to communicate a disposition of empathetic understanding in order to learn about someone else's social world and someone else's truth. Being considerate of people's views

and genuinely interested in listening to their perspectives has helped me gain their trust and develop a rapport with them. In the end, that kind of empathetic understanding is a major goal of ethnography.

What do you think sociologists gain from studying social life in the ways you do—through observation and in-depth interviews—that they can't get from other methods?

Hands down, the benefit of observation and in-depth interviews is deep understanding. I see myself as studying the realm of human activity where meaning and motivation reside. By that, I mean that I seek to understand the social meanings that guide the social actions of individuals and groups. Everyday life is messy and sometimes appears to be quite chaotic. There is tremendous variability in how people respond to situations, navigate institutions, and construct a life. Sociologists look to identify patterns—the underlying logic and routines within the chaos—through sustained and systematic observation and in-depth interviewing.

As a rule, the ethnographer attends to context in trying to understand patterns of social meaning and human behavior. When I read through my interview transcripts and ethnographic field notes, one question I always ask is, "What constraints are influencing how this person is behaving?" These can be constraints stemming from any number of things—organizational culture, formal or informal institutional rules, behavioral norms, obligation and belonging, exclusion, barriers of access, limits to opportunity, the structure of a particular situation. Our conduct depends on how situations are defined—and by whom. In watching what people do, I try to understand how societal constructs and constraints pattern our habits and our thinking.

A second strength of ethnography is the ability to understand social mechanisms that link social phenomena. Ethnography and other qualitative methods have lots of explanatory power. From deep immersion, we learn about the mechanisms—whether intrapersonal, interpersonal, situational, organizational, or institutional—that move people to act and produce specific outcomes.

As you've described in some of your writings on methodology (Best 2007), studying children can involve unique challenges. Do you have any advice in particular for aspiring sociologists who want to study more vulnerable populations? Have you come across situations when you were worried about how you were treating your research participants or portraying them in your writing? If so, how did you balance those concerns with your obligations to be an honest and rigorous researcher?

First, let me say I always feel worried about how I represent the worlds of the people I study. My advice to new researchers is you should be, too. Having that critical posture is essential if you're going to engage in honest and rigorous work. Ethnographers need to proceed cautiously, especially when we study across social or power differentials, or when we are outsiders to a setting—which is most of the time. When you study youth as an adult, for instance, it's really easy to get stuck in adult ways of knowing and impose that adult framework on young people's motivations and behavior. The same holds true if you're a white researcher studying across racial groups, or a middle-class researcher trying to understand the lives of people who aren't middle-class, but who are still judged by those standards. As a woman researcher, I have to recognize that gender is an ever-evolving social construct. I can't just draw upon my experience growing up as a girl to understand what it's like to be a girl now. I also have to consider the gender transformations of the last several decades.

There are ways to mitigate that worry and make it useful. First, you have to regularly check yourself. Question your personal assumptions, as well as your disciplinary, social, and cultural assumptions. Work to identify them, and assume you won't be able to fully access all of them.

Look for countervailing evidence—cases within your data that challenge or complicate the analysis you are developing, or what researchers call "negative cases."

Practice humility. When you are observing people, remember that you are not the expert on their life. They are. Your job is to listen and use the conceptual tools of sociology to understand their life in a new way. Engage in

"member checking"—where you seek confirmation from your participants that your analysis aligns with what is going on in their worlds. The goal should be to offer interpretations that generate empathetic understanding and allow actors to be seen in their full humanity.

Do your homework. There is a lot of excellent work already out there to serve as your guide. Build a diverse network of people who will keep you honest. When you plan your projects, build in time to step away from your work, and step back in. It will help you to see it from a different angle.

Another way to navigate some of these thorny issues about ethics and power imbalances in research is through the use of participatory research methods. In participatory research, the research participants are your co-researchers. In collaborative research, you partner with groups, which provide feedback at every step of the research process. These are two productive paths forward. While "member checking" remains a critical tool, participatory methods are more reciprocal, and participants have greater control over the outcomes of the knowledge generated from the research.

Ethnographers (and qualitative researchers more generally) have come under fire at times for overstepping ethical boundaries, including how we relate to our research participants and how we ensure the accuracy of our findings. Is there anything in particular you think we should be doing to address these concerns? What do you think of recent efforts to approach qualitative research in less conventional ways, such as making qualitative data public, publishing the real names of research subjects (with their permission), or conducting independent fact-checking of findings?

A lot of social science research is based on an extractive model. Ethnographers go into the field and usually gain more than they give. They largely maintain control of the final interpretation. That raises lots of prickly ethical questions, such as who benefits from our research. I find myself increasingly drawn to research methods that not only recognize the knowledge and expertise of the people being researched, but also have some direct benefit to them.

I'm also interested in the growing number of public repositories for qualitative data. That's not necessarily because I think we need more rigorous methods. Well-trained ethnographers committed to their craft are engaged in deep and robust inquiry. But what is exciting to me are the larger social gains made possible when data are publicly available. Some data, of course, will remain sensitive and should not be shared. But making systematically collected evidence widely obtainable will make it easier for scholars and the public to reach back into the past or access worlds different from their own. I think there is a lot of democratic possibility there.

In sociology, students are often encouraged to pursue quantitative methods over qualitative methods, partly because of fears about the professional and academic job markets that await them. What advice do you have for students who want to pursue qualitative research? How do you think these sorts of skills benefit them?

I think it's important to have basic statistical literacy for any job. At the same time, I find the conceptual and interpretive tools of qualitative sociology to be as important as the methodological tools of statistics. My sociology students have gone on to work in governmental and nongovernmental settings like nonprofits, as well as in the private sector. They tell me the broader conceptual tools they develop in sociology enable them to ask critical questions, evaluate evidence, and construct complex conceptual models that help them solve problems at multiple scales.

Key Takeaways

1. In approaching their fieldwork, some ethnographers are guided by grounded theory—generating theories purely inductively, from the ground up—while others take more hybrid approaches like the extended case method—refining existing theories through targeted observations.
2. When beginning an ethnographic study, carefully plan out where you will conduct your observations and what role you will adopt in the field. You should choose your field site guided by your overall theoretical approach—for instance, selecting a case that seems to contradict existing theories (per the extended case method). That said, also consider your personal flexibility in terms of time and location.
3. Like the participation continuum, researchers differ by degrees in the amount of disclosure they provide about their role as researchers while in the field. You should be aware of the trade-offs of standing at any point in this disclosure continuum. For instance, while covert research can give you unparalleled access to a group, it can pose ethical and logistical challenges given the need to misrepresent your identity and then later (when publication occurs) possibly out yourself as a researcher.
4. Ethnographers should continually engage in reflexivity, a process of self-reflection about how their own identity, beliefs, dispositions, actions, and practices may be influencing their research. There are some ascribed aspects of your identity that you cannot control, and sticking out within the group you are studying is not necessarily a bad thing. However, you will need to consider how you will go about gaining access and building rapport as an outsider, in addition to being careful about not misrepresenting or misinterpreting what you observe.

Exercises

1. Try to name at least three different locations where you might like to conduct field research. What barriers would you face were you to attempt to enter those sites as a researcher? In what ways might your entrée into the sites be facilitated by your positionality?
2. Watch **Video 9.2**, an interview with sociologist William Foote Whyte and a few of the children he studied (now grown men) in his ethnography *Street Corner Society*. What does the interview tell us about the advantages and disadvantages of ethnography as a research method? What does it tell you about the issues you need to be aware of when you interact with research participants in the field?
3. What is your opinion about researchers taking on a covert as compared with an overt role in field research? Which role would you like to take in an ethnographic study? Why?

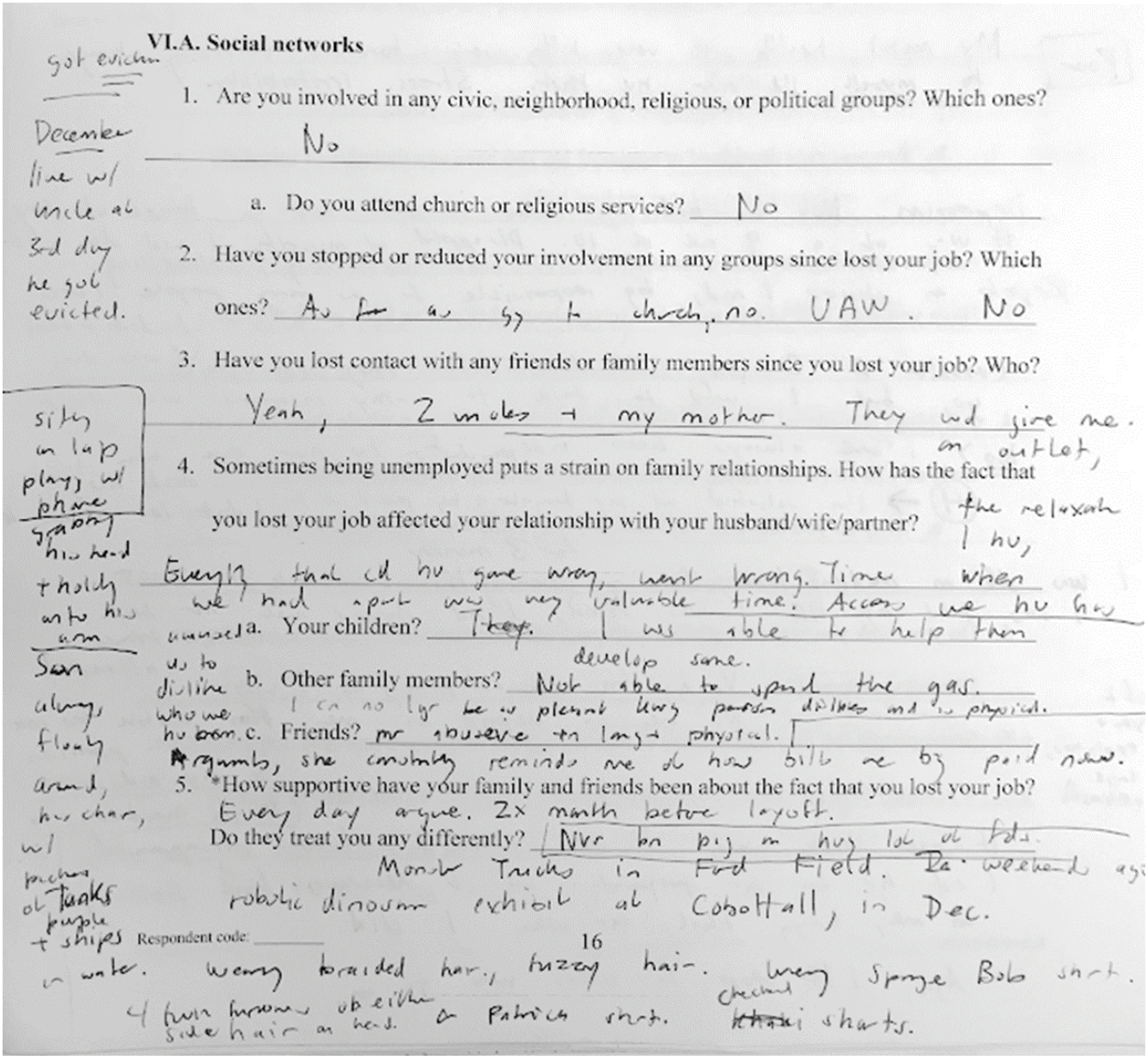
9.3. Field Jottings and Field Notes

Learning Objectives

1. Describe the various ways that ethnographers take notes while in the field.
2. Discuss what information is important to collect when taking field notes.

When ethnographers are observing in the field, they typically write off-the-cuff, descriptive notes about what they saw and heard. These are called **field jottings** (Figure 9.4). Then, when they leave the field, they take these jottings and put together a memo of their observations for that particular time period, which is known as a **field note**. Let's go over each type of note-taking, both of which are essential to the ethnographic process.

Field Jottings: Writing in the Field



got evicted

December live w/ uncle at 3rd day he got evicted.

VI.A. Social networks

1. Are you involved in any civic, neighborhood, religious, or political groups? Which ones?
   No
   a. Do you attend church or religious services? No
2. Have you stopped or reduced your involvement in any groups since lost your job? Which ones? As far as go to church, no. UAW. No
3. Have you lost contact with any friends or family members since you lost your job? Who?
   Yeah, 2 mos + my mother. They wd give me. an outlet,
4. Sometimes being unemployed puts a strain on family relationships. How has the fact that you lost your job affected your relationship with your husband/wife/partner? the relaxation I hv,
   Everything that cd hv gone wrong, went wrong. Time when we had spent w/ very valuable time. Accom we hv hv
   a. Your children? They. I was able to help them develop some.
   b. Other family members? Not able to spend the gas.
   c. Friends? mr abusive to long + physical. [illegible]
5. *How supportive have your family and friends been about the fact that you lost your job?
   Every day argue. 2x month before layoff.
   Do they treat you any differently? Nvr bn pig in hvy lot of fds.
   Monster Trucks in Ford Field. Da weekend ago

sitting in lap plays w/ phone grabs his head holds onto his arm

Sean always floats around, has char, w/ packs of Funks purple + ships in water.

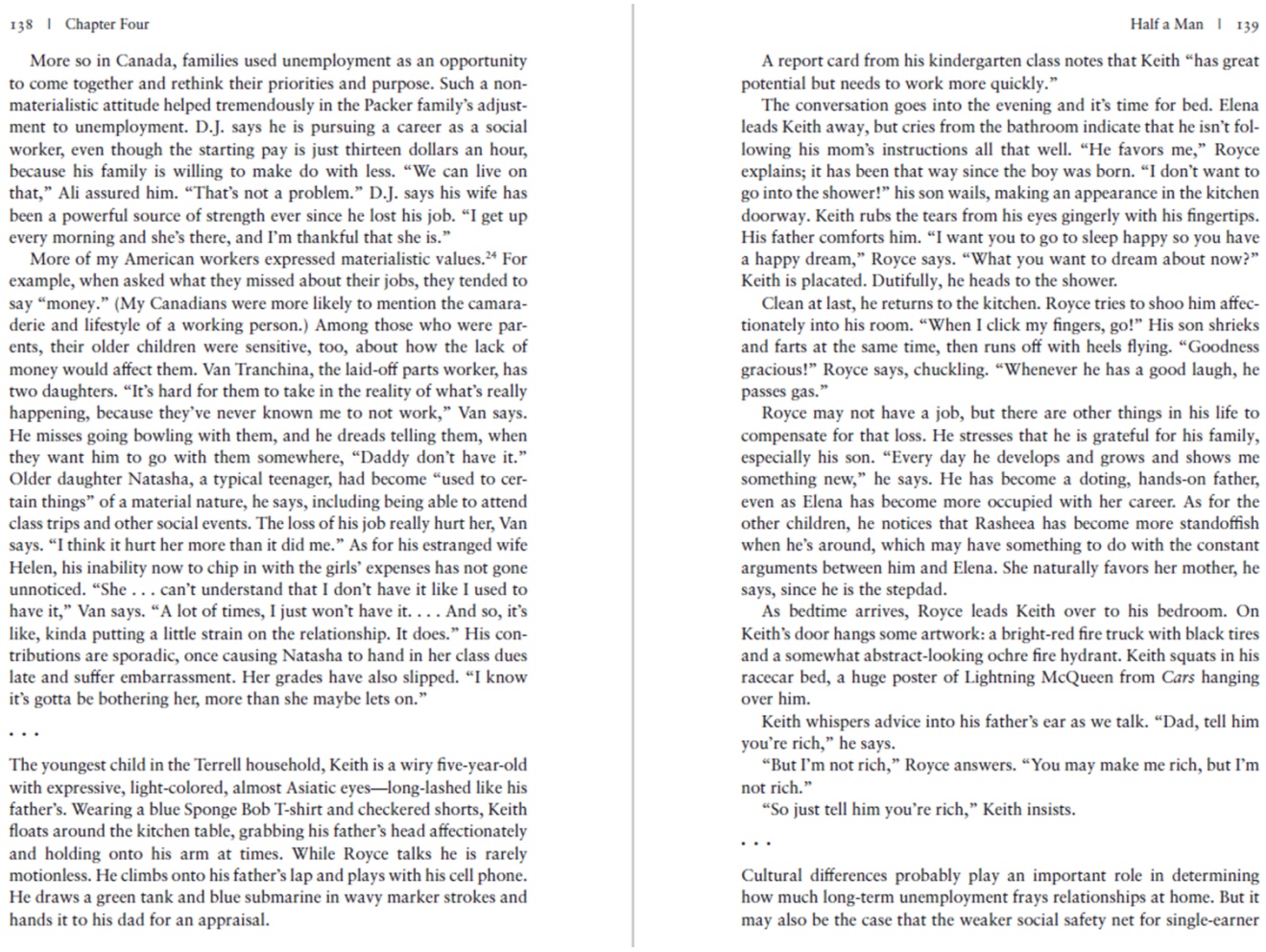
robotic dinosaurs exhibit at CobbHall, in Dec.

Respondent code: _____ 16

wearing braided hair, fuzzy hair. wearing Sponge Bob shirt.

4 twin brothers ob alike side hair on heads. a Patrick shirt. checked shirts.

Figure 9.4a. Field Jottings. You can write your field jottings anywhere—in a notebook, in your cellphone's notes app, on your hand. For the fieldwork used in his book *Cut Loose* (2015), Victor often wrote down observational notes (shown in this first image) about his participants in the margins of the interview guide he used for his conversations with unemployed workers.



138 | Chapter Four

More so in Canada, families used unemployment as an opportunity to come together and rethink their priorities and purpose. Such a nonmaterialistic attitude helped tremendously in the Packer family's adjustment to unemployment. D.J. says he is pursuing a career as a social worker, even though the starting pay is just thirteen dollars an hour, because his family is willing to make do with less. "We can live on that," Ali assured him. "That's not a problem." D.J. says his wife has been a powerful source of strength ever since he lost his job. "I get up every morning and she's there, and I'm thankful that she is."

More of my American workers expressed materialistic values.[24] For example, when asked what they missed about their jobs, they tended to say "money." (My Canadians were more likely to mention the camaraderie and lifestyle of a working person.) Among those who were parents, their older children were sensitive, too, about how the lack of money would affect them. Van Tranchina, the laid-off parts worker, has two daughters. "It's hard for them to take in the reality of what's really happening, because they've never known me to not work," Van says. He misses going bowling with them, and he dreads telling them, when they want him to go with them somewhere, "Daddy don't have it." Older daughter Natasha, a typical teenager, had become "used to certain things" of a material nature, he says, including being able to attend class trips and other social events. The loss of his job really hurt her, Van says. "I think it hurt her more than it did me." As for his estranged wife Helen, his inability now to chip in with the girls' expenses has not gone unnoticed. "She . . . can't understand that I don't have it like I used to have it," Van says. "A lot of times, I just won't have it. . . . And so, it's like, kinda putting a little strain on the relationship. It does." His contributions are sporadic, once causing Natasha to hand in her class dues late and suffer embarrassment. Her grades have also slipped. "I know it's gotta be bothering her, more than she maybe lets on."

. . .

The youngest child in the Terrell household, Keith is a wiry five-year-old with expressive, light-colored, almost Asiatic eyes—long-lashed like his father's. Wearing a blue Sponge Bob T-shirt and checkered shorts, Keith floats around the kitchen table, grabbing his father's head affectionately and holding onto his arm at times. While Royce talks he is rarely motionless. He climbs onto his father's lap and plays with his cell phone. He draws a green tank and blue submarine in wavy marker strokes and hands it to his dad for an appraisal.

Half a Man | 139

A report card from his kindergarten class notes that Keith "has great potential but needs to work more quickly."

The conversation goes into the evening and it's time for bed. Elena leads Keith away, but cries from the bathroom indicate that he isn't following his mom's instructions all that well. "He favors me," Royce explains; it has been that way since the boy was born. "I don't want to go into the shower!" his son wails, making an appearance in the kitchen doorway. Keith rubs the tears from his eyes gingerly with his fingertips. His father comforts him. "I want you to go to sleep happy so you have a happy dream," Royce says. "What you want to dream about now?" Keith is placated. Dutifully, he heads to the shower.

Clean at last, he returns to the kitchen. Royce tries to shoo him affectionately into his room. "When I click my fingers, go!" His son shrieks and farts at the same time, then runs off with heels flying. "Goodness gracious!" Royce says, chuckling. "Whenever he has a good laugh, he passes gas."

Royce may not have a job, but there are other things in his life to compensate for that loss. He stresses that he is grateful for his family, especially his son. "Every day he develops and grows and shows me something new," he says. He has become a doting, hands-on father, even as Elena has become more occupied with her career. As for the other children, he notices that Rasheea has become more standoffish when he's around, which may have something to do with the constant arguments between him and Elena. She naturally favors her mother, he says, since he is the stepdad.

As bedtime arrives, Royce leads Keith over to his bedroom. On Keith's door hangs some artwork: a bright-red fire truck with black tires and a somewhat abstract-looking ochre fire hydrant. Keith squats in his racecar bed, a huge poster of Lightning McQueen from *Cars* hanging over him.

Keith whispers advice into his father's ear as we talk. "Dad, tell him you're rich," he says.

"But I'm not rich," Royce answers. "You may make me rich, but I'm not rich."

"So just tell him you're rich," Keith insists.

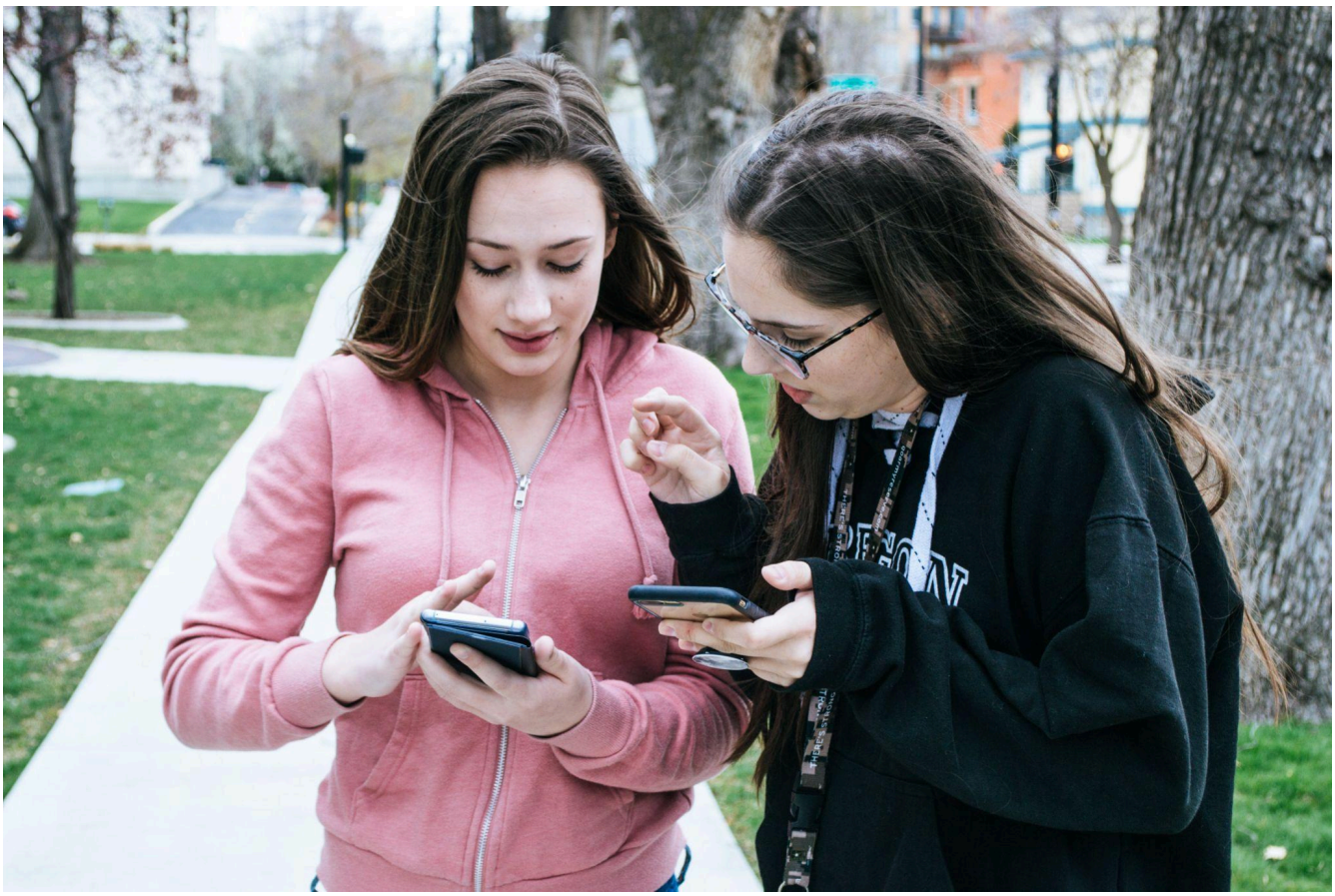
. . .

Cultural differences probably play an important role in determining how much long-term unemployment frays relationships at home. But it may also be the case that the weaker social safety net for single-earner

Figure 9.4b. Vignette. In this second image, you can see how these field jottings eventually turned into a vignette included in the published book (V. T. Chen 2015:138–139). *University of California Press*

As an ethnographer, there isn't one right way to write field jottings. If you can conduct your fieldwork in a setting where sitting with a notebook, tablet, or computer is no problem—such as when you are observing a classroom or meeting—then recording your observations at length and in detail will be pretty straightforward. If you are in a location where taking notes would be rude or make people uncomfortable, however, you have to weigh the benefits of extensive note-taking against the dangers of sticking out. Covert researchers will obviously not want to be seen taking notes, which may prompt questions that blow their cover. Yet even overt researchers should be concerned about how people will act differently when they are aware that someone is writing down everything they do or say. A lack of discretion can open up the possibility of social desirability bias or other reactivity problems of the sort that we described earlier.

Your strategy for recording your observations while in the field will be determined mostly by the site you choose and the role you decide to play in that site. Will you be in a setting where having a notebook or smartphone in your hands will look out of place? If not, by all means, take extensive notes! But don't let your note-taking distract you from what's happening around you. At times, you should just take in the scene. If you are a participant in addition to an observer, you may also want or need to focus on engaging with the setting and the people within it. In fact, participant observers often find that they have their hands too full to write any notes at all (more about that later).



People are on their cellphones always and everywhere—which means if you want to write down field jottings in an inconspicuous manner, you can just pretend to be texting on your phone. *Brett Sayles, via Pexels*

When copious note-taking isn't possible or desirable, you can record your observations in more casual ways. For instance, the fact that smartphones are so common nowadays means that ethnographers can unobtrusively write down their observations in a notes app or email while appearing, like everyone else, to be texting or surfing the web. Other strategies can include keeping a tiny notebook in your pocket and whipping it out whenever you need to jot something down, dictating quick observations into your phone's voice memo app or a recorder, or—if you really want to go full-on Cold War spy—writing key phrases on your hands and arms or scribbling notes on toilet paper during visits to the restroom. With these more surreptitious approaches to note-taking, you'll need to develop a way of jotting down observations that doesn't require complete sentences or perhaps even words. Sometimes, ethnographers develop their own version of shorthand to take notes, using some combination of abbreviations and symbols. Other times, they will take pictures or videos of what is going on around them, which they can refer to later to jog their memories of what they observed. (Be aware of your IRB's stance toward the collection of such supplementary data, however.)

At the jottings stage, the notes you write will be descriptive. That is, you will want to describe what you saw or heard (and, if applicable, smelled, tasted, and touched) as clearly and deeply as possible. Draw pictures or diagrams if it's helpful. You might also want to make note of some initial personal impressions or reactions (more about this later), though generally speaking, you will want to provide a straightforward, factual account. Remember that your ultimate goal is thick description, a detail-rich account of what you observed that connects people's actions and words to their

specific context. While the cultural interpretations and layered social meanings of your observation won't necessarily be clear at this stage, by writing up a vivid description you will be in a good position later on to engage in that sort of analysis.

Here are some questions you might ask yourself in the field, adapted from a helpful list developed by Vanderbilt University's Writing Studio (2012):

Space:

1. What is the layout of the space or room?
2. What are the specific objects or physical elements in the space?
3. Do these features of the space affect the interactions taking place in any way? Do they communicate anything about the people frequently in this space?

People:

1. Who are the people involved? (Single out specific individuals and interactions to describe, in addition to providing an overall sense of the composition and tendencies of the group or groups.)
2. What are people wearing? What does it say about who they are, or how they are feeling in the moment?
3. What clues signify people's statuses and roles? Are there aspects of their appearance or demeanor that signal how they relate to other people in this space?

Actions and motivations:

1. What are the people you are observing doing in general or attempting to accomplish?
2. What are individuals' specific behaviors—both verbal and nonverbal?
3. How do people interact with one another? How do they interact with their surroundings? (Single out specific interactions to describe, in addition to providing an overall sense of how the group or groups interact.)

Emotions:

1. What is the affect of the people you are observing? What emotion (or noticeable lack of emotion) are they expressing? How are they expressing it?
2. What clues are there about people's feelings about one another or the nature of their relationships?

Rules and norms:

1. What structures, rules, or norms govern the situation? How explicit or implicit are they?
2. How are expectations conveyed? Are there individuals who are policing these rules or norms? Are there posted signs or other nonverbal ways of communicating what is expected?

Writing useful field jottings is a skill that you will need to master. The observation involved here is deliberate, not haphazard—though it may seem haphazard to a novice researcher. With practice, you'll get a better sense of when to write, what to write, where to write, and how to write. For instance, try observing a setting in your neighborhood, school, or workplace for just 15 minutes. How hard can it be? Pretty tough, if you do it right. There is always too much detail to write down—everything from how people dress, to what the weather is like that day, to how the space is laid out, to what expressions and tone of the voice people use while interacting. This becomes all the more apparent if you compare your notes with those of other people who observed the exact same setting. (In **Exercise 1** at the end of this section, we'll run a little experiment along these lines.) There may be overlap in the things each of you note,

but inevitably one person will have paid more attention to conversations overheard, another person to people's actions and mannerisms, and yet another to the nonhuman dimension—the surrounding landscape, the ambient sounds and scents, and so on. You and your fellow observers will probably use different note-taking strategies, too: some might write complete sentences, while others might have abbreviations, idiosyncratic personal symbols, or drawings in their notes.

In conducting this sort of exercise, you're likely to experience not just how difficult the task of observation can be, but also how arbitrary and aimless it can feel. Having a research question or topic in mind helps a researcher focus their observations, and that is one lesson to learn: come into the field with some plan about what you want to observe, so that you aren't overwhelmed by the sheer amount of data. That said, part of the point of ethnography's inductive approach is to make sense of the seemingly chaotic interactions and activities witnessed in the field. In that sense, the ethnographer needs to be comfortable with a certain degree of ambiguity, and with making their own decisions about what aspects of the vast array of phenomena they observe is worth jotting down. As usual, there is not often a right or best answer. Furthermore, it is important that ethnographers not allow their original question or topic blind them to other occurrences in the field. When conducting observations, you want to write down as much as you can, because oftentimes something that may seem unimportant at the time ends up being important down the line—sometimes only years later, when you're writing up your work for publication. As we've emphasized, one of the strengths of ethnography is its ability to generate new theories, rather than just refining or testing existing ones, and that creativity depends on observing the unexpected.



In the field of journalism, there is a saying that if a reporter didn't write down something in their notebook, it didn't happen. Much the same thing can be said for ethnography: if you observe something but neglect to note it, it might as well not have happened. *Tima Miroshnichenko, via Pexels*

No matter how difficult it can be to write jottings while in the field, it is worth the effort. Ethnographers rely on them to develop more complete field notes later on. Even if you have to be extremely cautious or spare in your note-taking, try to jot down a few things very quickly while you're in the field. Something is better than nothing. You may have the temptation to skip the jottings stage and just write down what you remember in a field note, but this is a risky strategy. Your memory will quickly fade once you leave the field, and the amount of detail you can recall just hours later will be much less than what you can reconstruct from jottings. Furthermore, research finds that eyewitness accounts actually become distorted with time, as people unconsciously revise the memories in their minds based on what is most familiar or expected.

As we pointed out earlier, sometimes writing jottings in the field is utterly not an option—in which case you will want to write down everything right after your observation. But, in general, both field jottings and field notes (the more formal note-taking we'll describe in the next section) are critical. Together, they allow the researcher to generate the thick description that allows for insightful analysis.

Field Notes: Processing What You Observed



Steve Viscelli (2016) learned to drive a big rig and did a paid stint running cargo across the country in order to understand the brutally competitive and exploitative industry of long-haul trucking. After a long day on the road, Viscelli said he would be "dead tired," but he would still record notes in a personal journal, sometimes in audio form. *World Sikh Organization of Canada, via Pexels*

The next stage in the ethnographic process can begin as soon as you end your current observation session, your messy jottings or recordings in hand (in some cases, literally *on* hand), At some point, you'll want to sit at a computer to type up those notes into a more readable format. That said, we recommend that you take some time to review and possibly flesh out your field jottings immediately upon leaving any observation in the field, even if you can't turn them into a formal field note just yet.

When you finally start writing your field note, you typically will want to create a new document in your word processor devoted to that single observation session. One alternative is to write your field notes as dated entries in a single document, but we'd suggest having each field note be its own document to make your life easier later on. As shown in **Figures 9.5a** and **9.5b**, this approach allows you to order your field notes by date and store them in folders representing different types of observations (e.g., what organization, site, or event the observation focused on). As you'll see in **Chapter 11: Qualitative Data Analysis**, storing your field notes in separate files also works well with the qualitative data analysis software packages you might use during the coding process, which naturally treat each document as a separate observation.



Figure 9.5a. Organizing Field Notes by Event or Organization. It is essential to find a good organizing scheme for all the data you will collect during an ethnographic study. This first screenshot shows how you might categorize your field notes by event type or organization.

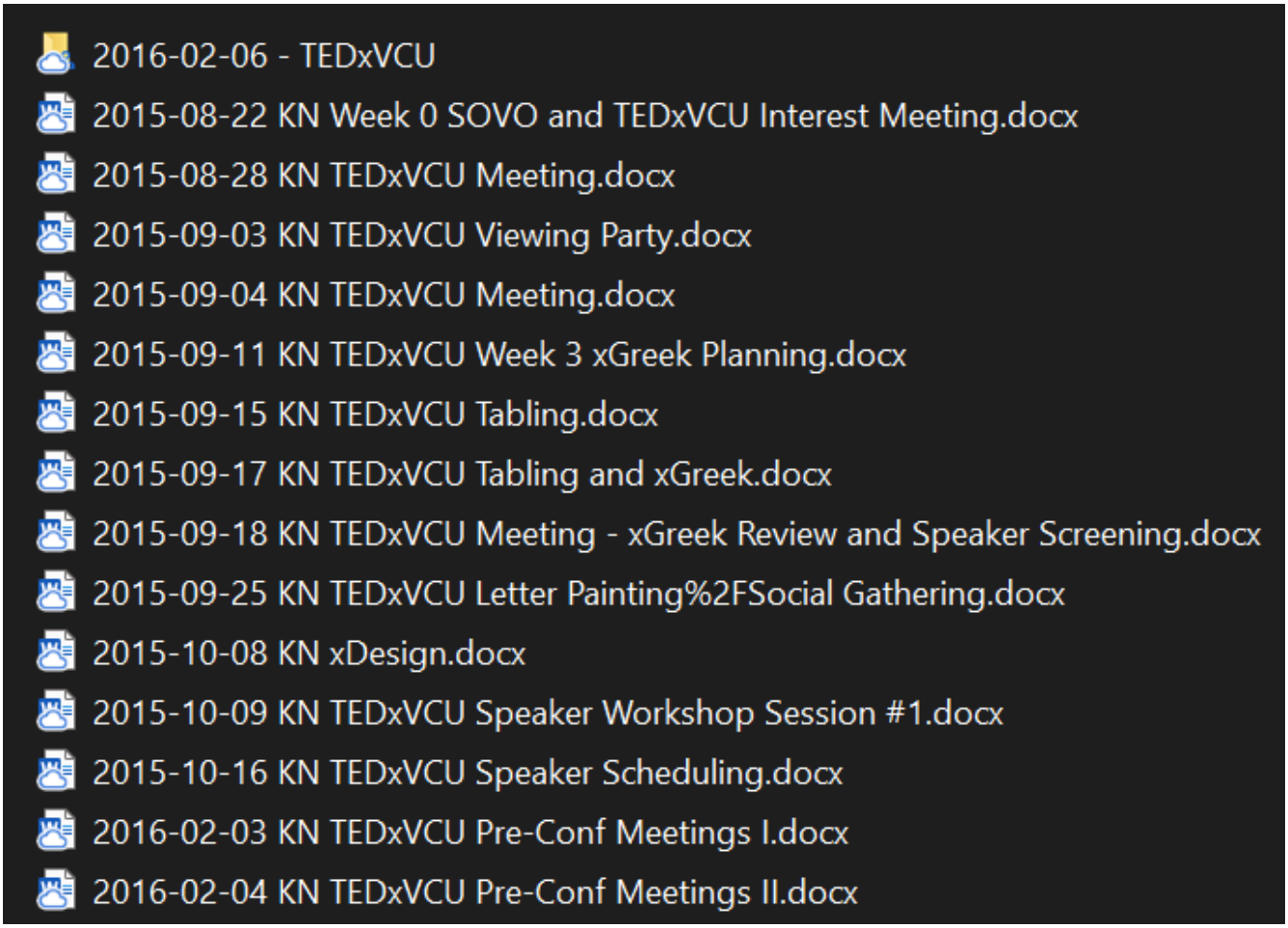


Figure 9.5b. Organizing Field Notes by Date. This second screenshot shows how you might order your field notes by date. Note that starting the names of your files with a year/month/day format like YYYY-MM-DD will order them chronologically by default in their folder.

As you write up your field note, don't worry so much about the quality of your prose. Sometimes ethnographers quote directly from their field notes in their published work, but that is not required, and we'd suggest not overthinking these memos: just write whatever comes to mind, as soon afterward as you can. The goal is to record your observations as straightforwardly and quickly as possible in a way that makes sense to you. Even if you felt that the descriptions you recorded in the field were more or less complete, you'll be surprised by how much more you're able to recall once you sit down without distractions and read through what you've jotted down. You'll also have the opportunity to add your own reflections—your observations about your observations—when you reach this stage in the research process.

In your field note, fill in the blanks of your jottings, writing down as much as possible about what you've observed. Even if the details seem microscopic or mundane, they may come in handy later. (Minute details may also make your narratives a richer and more compelling read, helping your sociological work reach a broader audience.) Indeed, some say field notes can never contain too much detail—though you also want to be kind to yourself as a researcher and not make every brief episode of observation into such an arduous note-taking task that you are discouraged from keeping up with your observations (remember, ethnography is a marathon, not a sprint).

Ethnographers typically spend several hours typing up field notes after each observation has occurred. A good rule of thumb is that every hour of observation requires at least two hours writing up field notes. You should pace yourself and figure out what note-writing approach works best for you. For example, for his ethnography of drug robbers in the South Bronx, Randol Contreras began his project writing up meticulous field notes about his interactions and observations. After three months, however, he started putting a tape recorder in his jacket or pants pocket and bringing it everywhere he went. "The tape recorder improved my recollection of events since I often got back home in the early morning, usually intoxicated from heavy drinking," Contreras writes (2012:13). "Nevertheless, I wrote up extensive outlines before going to bed (which sometimes took over two hours) and then wrote elaborate field notes the next day. These notes supplemented and guided the tape recordings, especially since street sounds sometimes interfered with the sound quality, and sometimes weeks passed before I found time for transcription."

Writing as much as possible, in as much detail as possible, should also help you avoid generalizing in your field notes. As in good fiction writing, the mantra here is "show, don't tell." Be specific about what you observe. You won't end up using all or even most of the details you're writing down, but when you get to the analysis and writing stages of your research, you want as much material as possible. It will help trigger ideas for nuanced and well-grounded theories, and it will allow you to immerse the reader in the social contexts you observed.

As outlined in **Table 9.2**, avoid any vague descriptions and generic adjectives in your write-up. Rather than saying that someone you observed was "angry," describe what gave you that impression—was that person yelling, red in the face, shaking their fist? Rather than describing people as "nice" and "fashionable" or a setting as "fancy," say what exactly a child did to appear "nice," what specific items of clothing a woman wore that made her seem "fashionable," and what kinds of decor, lighting, music, and even table settings made a restaurant seem "fancy." Rather than saying that "everyone" said or did something, make note of exactly *who* said or did X (or note that you're not sure exactly who did so, but that it *seemed* as if most everyone did).

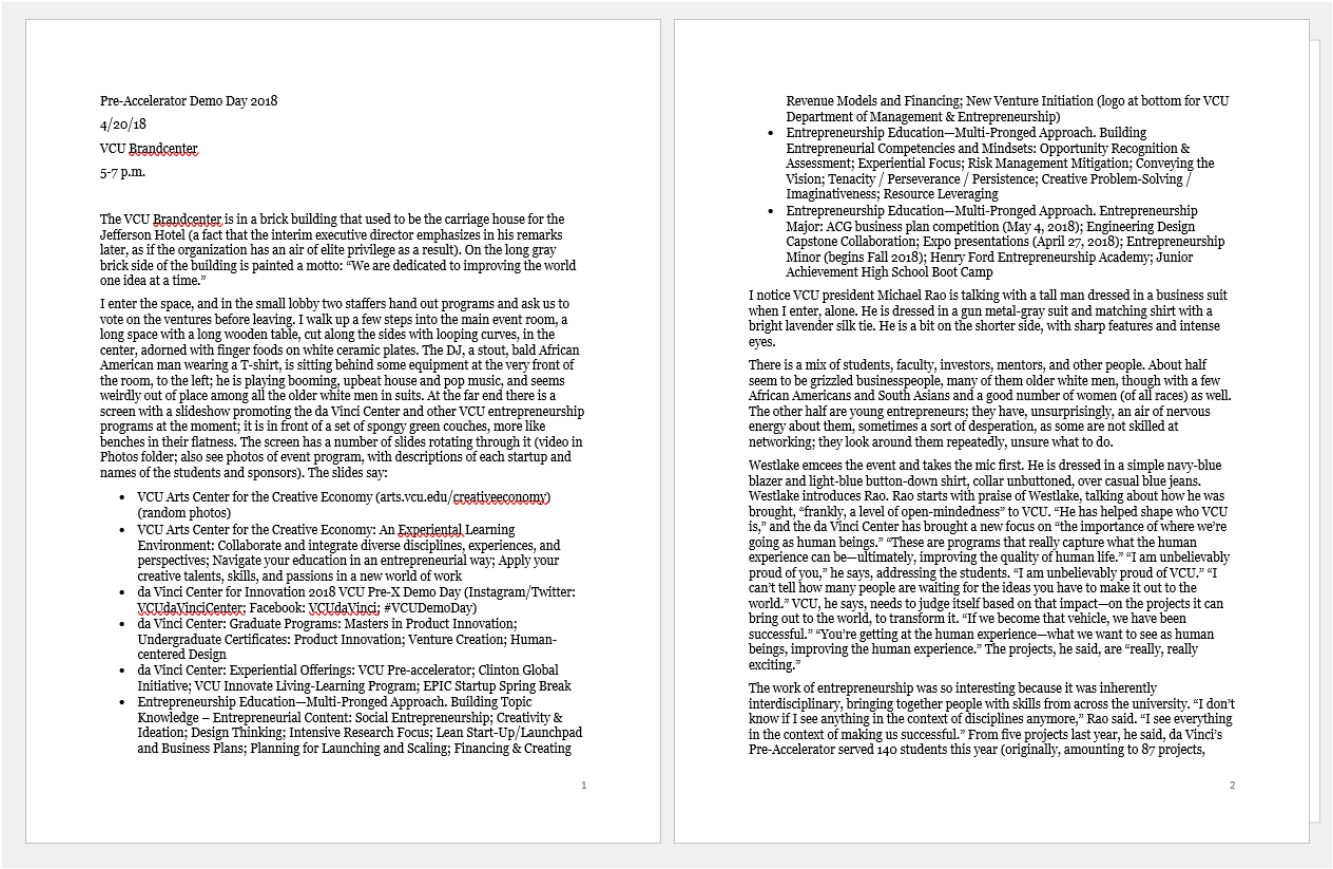
Use similes and analogies to help make your characterizations more concrete and visceral—for instance, describing a "large" room as the size of a football field, or comparing an adult's "violent" outburst to a red-faced toddler's tantrum. Capture specific interactions you observed—noting how you participated, if you did. And summarize conversations you overheard, providing direct quotes whenever possible. Even a few words of remembered dialogue can say so much more than a blanket description of how a person expressed disapproval or happiness.

Table 9.2. Telling versus Showing

Telling	**Showing**
"I'm sorry," she said, clearly upset.	"I'm sorry," she said, looking down at the ground with slumped shoulders.
He was dressed sloppily.	He wore a red Henley shirt over faded light-blue jeans with a hole ripped into the right knee. His shirt was untucked and none of the three buttons were buttoned.
The protesters seemed aimless and confused.	The protesters wandered through the hallways, popping their heads into open doorways. Those at the front of the crowd seemed to lack any direction or plan, moving hesitantly through the building like the blind leading the blind.
"We love you!" the crowd shouted, cheering her on.	"We love you!" shouted a middle-aged man from the upper balcony, and the rest of the crowd started whooping as the singer picked up the microphone.

Don't forget to describe exactly where you were and detail your surroundings. Physical descriptions of the people you observe can also be helpful, but throughout all of this data collection, be conscious about your IRB's strictures regarding confidentiality. When describing people or objects physically, one trick is to go from top to bottom (or left to right) making note of every detail you see along the way. This has the added benefit of focusing your attention on what you can concretely show ("he had droopy eyelids") rather than vaguely tell ("he was tired").

Early in an ethnographic study, you will probably focus more on describing your surroundings, the appearance of the people you encounter, and other introductory details than you do later on, when the focus will be primarily on how people interact or what events transpire. This means that your initial field notes will include very detailed descriptions of the locations you observe and the people with whom you interact; you might also draw a map, or—if appropriate in your setting—take pictures or videos of your field sites. It is crucial to have documentation early on of your first impressions, and with time, these sorts of details may become so much a part of the everyday scene that you stop noticing them. If your observations are being conducted in the same places and with the same people, you won't need to write up the same descriptions in later field notes; instead, you can focus on the interactions and events you observe, updating your initial descriptions of individuals or surroundings only as they change in relevant ways.



Pre-Accelerator Demo Day 2018

4/20/18

VCU Brandcenter

5-7 p.m.

The VCU Brandcenter is in a brick building that used to be the carriage house for the Jefferson Hotel (a fact that the interim executive director emphasizes in his remarks later, as if the organization has an air of elite privilege as a result). On the long gray brick side of the building is painted a motto: "We are dedicated to improving the world one idea at a time."

I enter the space, and in the small lobby two staffers hand out programs and ask us to vote on the ventures before leaving. I walk up a few steps into the main event room, a long space with a long wooden table, cut along the sides with looping curves, in the center, adorned with finger foods on white ceramic plates. The DJ, a stout, bald African American man wearing a T-shirt, is sitting behind some equipment at the very front of the room, to the left; he is playing booming, upbeat house and pop music, and seems weirdly out of place among all the older white men in suits. At the far end there is a screen with a slideshow promoting the da Vinci Center and other VCU entrepreneurship programs at the moment; it is in front of a set of spongy green couches, more like benches in their flatness. The screen has a number of slides rotating through it (video in Photos folder; also see photos of event program, with descriptions of each startup and names of the students and sponsors). The slides say:

- VCU Arts Center for the Creative Economy (arts.vcu.edu/creativeeconomy) (random photos)
- VCU Arts Center for the Creative Economy: An Experiential Learning Environment: Collaborate and integrate diverse disciplines, experiences, and perspectives; Navigate your education in an entrepreneurial way; Apply your creative talents, skills, and passions in a new world of work
- da Vinci Center for Innovation 2018 VCU Pre-X Demo Day (Instagram/Twitter: VCUdaVinciCenter; Facebook: VCUdaVinci; #VCUDemoDay)
- da Vinci Center: Graduate Programs: Masters in Product Innovation; Undergraduate Certificates: Product Innovation; Venture Creation; Human-centered Design
- da Vinci Center: Experiential Offerings: VCU Pre-accelerator; Clinton Global Initiative; VCU Innovate Living-Learning Program; EPIC Startup Spring Break
- Entrepreneurship Education—Multi-Pronged Approach. Building Topic Knowledge – Entrepreneurial Content: Social Entrepreneurship; Creativity & Ideation; Design Thinking; Intensive Research Focus; Lean Start-Up/Launchpad and Business Plans; Planning for Launching and Scaling; Financing & Creating Revenue Models and Financing; New Venture Initiation (logo at bottom for VCU Department of Management & Entrepreneurship)
- Entrepreneurship Education—Multi-Pronged Approach. Building Entrepreneurial Competencies and Mindsets: Opportunity Recognition & Assessment; Experiential Focus; Risk Management Mitigation; Conveying the Vision; Tenacity / Perseverance / Persistence; Creative Problem-Solving / Imaginativeness; Resource Leveraging
- Entrepreneurship Education—Multi-Pronged Approach. Entrepreneurship Major: ACG business plan competition (May 4, 2018); Engineering Design Capstone Collaboration; Expo presentations (April 27, 2018); Entrepreneurship Minor (begins Fall 2018); Henry Ford Entrepreneurship Academy; Junior Achievement High School Boot Camp

I notice VCU president Michael Rao is talking with a tall man dressed in a business suit when I enter, alone. He is dressed in a gun metal-gray suit and matching shirt with a bright lavender silk tie. He is a bit on the shorter side, with sharp features and intense eyes.

There is a mix of students, faculty, investors, mentors, and other people. About half seem to be grizzled businesspeople, many of them older white men, though with a few African Americans and South Asians and a good number of women (of all races) as well. The other half are young entrepreneurs; they have, unsurprisingly, an air of nervous energy about them, sometimes a sort of desperation, as some are not skilled at networking; they look around them repeatedly, unsure what to do.

Westlake emcees the event and takes the mic first. He is dressed in a simple navy-blue blazer and light-blue button-down shirt, collar unbuttoned, over casual blue jeans. Westlake introduces Rao. Rao starts with praise of Westlake, talking about how he was brought, "frankly, a level of open-mindedness" to VCU. "He has helped shape who VCU is," and the da Vinci Center has brought a new focus on "the importance of where we're going as human beings." "These are programs that really capture what the human experience can be—ultimately, improving the quality of human life." "I am unbelievably proud of you," he says, addressing the students. "I am unbelievably proud of VCU." "I can't tell how many people are waiting for the ideas you have to make it out to the world." VCU, he says, needs to judge itself based on that impact—on the projects it can bring out to the world, to transform it. "If we become that vehicle, we have been successful." "You're getting at the human experience—what we want to see as human beings, improving the human experience." The projects, he said, are "really, really exciting."

The work of entrepreneurship was so interesting because it was inherently interdisciplinary, bringing together people with skills from across the university. "I don't know if I see anything in the context of disciplines anymore," Rao said. "I see everything in the context of making us successful." From five projects last year, he said, da Vinci's Pre-Accelerator served 140 students this year (originally, amounting to 87 projects,

1

2

Figure 9.6. Example of a Field Note. As part of a study of entrepreneurship (Chen and Goldstein 2024), Victor wrote this field note after attending an event where Virginia Commonwealth University students give public pitches about their startup ideas.

Figure 9.6 offers an example of a field note from Victor's research. One thing you'll notice is that Victor used quotation marks every time a person was directly quoted. He also paraphrased people's words without quotes (with an attribution like "said") when he wasn't confident he captured the exact quote. In your own field notes, include as many quotes

as you can. In particular, direct quotes can provide important support for the analytic points you'll make when you later describe patterns in your data. This is another reason that taking notes in the field is a good idea: direct quotes may be impossible to remember hours or even minutes after hearing them. (For this reason you may also wish to focus in your field jottings on writing out any pertinent verbatim quotes and then only later take the time to describe the circumstances under which something was said when you write up your field notes.) That said, partial quotes or paraphrases are fine if you can't catch people's precise wording. Again, you are never going to capture everything, so do your best, pace yourself, and don't beat yourself up if you miss something you think is important—there will be more opportunities to follow up.

Another thing you can see in the example field note in **Figure 9.6** is that researchers will record their own feelings and impressions in their field notes. The personal reflections in your field notes may never be used, but in some cases, they will serve as a preliminary analysis of your data. At times, researchers will also use their field notes as a constructive opportunity to "vent"—express emotion or purge difficult thoughts or feelings that arise when they become upset about or annoyed by something that occurred in the field. As we've noted, ethnography requires developing personal relationships with participants, and any interpersonal relationship can become fraught or volatile. Feel free to express your feelings in your notes. These details may become important for analysis later, but even if not, writing them down can be cathartic. They might also reveal biases you have about the participants that you should confront and be honest about.

Every researcher's approach to writing up field notes will vary according to whatever strategy works best for that individual. For example, some ethnographers will distinguish between statements in their notes that are purely descriptive and those that are personal reflections or preliminary analyses. They may segregate the two kinds of content in separate sections of a field note. They may create two columns—one containing notes only about what was observed directly, and the other containing reactions and impressions. Or they may flag the latter commentary in special ways—adding brackets around the text, putting it in a different font or color, or embedding it in footnotes or comments. You don't need to do all of this, however, and you should do whatever works best for you personally. What's important is that you adopt a strategy that enables you to write accurately, record as much detail as possible, and distinguish observations from reflections.

Oftentimes, the tenor of your field notes will change as you conduct more observations. You may find yourself shifting from doing a lot of describing to doing much more explaining in your write-ups. In fact, some ethnographers say that there are two different kinds of field notes: *descriptive* and *analytic*. **Descriptive field notes** are notes that simply describe a researcher's observations as straightforwardly as possible; **analytic field notes** attempt to explain or comment on those observations. Of course, the line between what counts as "description" and what counts as "analysis" can get pretty fuzzy, and you may prefer not to compartmentalize your thinking in this way, but just let your field notes naturally take a more explanatory turn over time. Even so, one technique you might consider is to append an explicitly analytic field note to the end of the earlier, more descriptive one written immediately following your observations. This gives you the option at any later point to go back to that initial write-up and expound on your initial impressions and whatever thoughts have come to you since then.

With all the time you spend writing up field notes, you may be wondering why you aren't just writing your paper manuscript at this point. Why take this extra step between field jottings and analysis? Field notes are the foundation of a quality analysis. As we said, you will find that you will be able to remember much more when you sit down to write up your jottings in a formal memo. Furthermore, you will start drawing interesting connections and thinking through how exactly the relevant social context of what you observed matters in terms of what people did or said. Once you move to the formal analysis stage, which we describe in **Chapter 11: Qualitative Data Analysis**, you will be able to draw from all these memos—which will number dozens or hundreds of documents at this point—and see the patterns across your observations from a higher-level vantage point. (In fact, as we'll discuss in that later chapter, you can even import your field notes into a computer program, treating them as separate sources of data that you can easily categorize and

compare.) All these stages of ethnographic research contribute significantly to the insightfulness and rigor of the work you ultimately produce. Finally, it's worth noting that your field notes are the hard evidence and formal record of what you observed. If another researcher wanted to vet your ethnography, they'd look to your field notes to get a sense of your data and how you drew the conclusions you did. In fact, many ethnographers count their field notes or even tally up all the pages of written text to show that they did their due diligence in the field—for instance, saying something like their observations generated "100 pages" of field notes. (Alternatively, they will talk about how many events they observed, how many hours they spent in the field, and so on.) Although we think it is generally a bad idea to fixate on quantifying your qualitative fieldwork—remember, depth rather than breadth is the key advantage of ethnography—you will want to keep some sort of accounting of the scale of your observations in case other scholars want more in the way of proof regarding the rigor of your ethnography.

Key Takeaways

1. When taking notes in the field, try to record your observations as clearly and deeply as possible. It may be necessary to find unobtrusive ways of taking notes in the field, such as typing out notes on your smartphone.
2. Note-taking does not end when a researcher exits the field. Try to type up your jottings as extended field notes immediately upon leaving the field so that you can expand upon what you observed (and what you personally felt in reaction to those observations) before your memory fades. Whenever possible, including analytic sections in your field notes that provide a preliminary analysis of puzzling occurrences or interesting patterns, which can help influence your next round of observations in the field.
3. While observing, capture as much information as possible using all your senses. Describe individuals, surroundings, and interactions. Single out specific examples in addition to noting larger patterns. Details that may seem unimportant in the moment may turn out to be important during later analysis.
4. Remember to show, not tell, when describing scenes or individuals. This generates rich detail that provides a firmer foundation for theorizing—in addition to helping you immerse readers in the social world you observed.

Exercises

1. Recruit a friend or two to accompany you to a public space—perhaps a library, cafeteria, coffeeshop, or park. Each of you should observe the space and the people within it for about 15 minutes., writing notes to capture as many details as you can. It's best to sit or stand near your friends, but do not talk with them so that you don't influence what the other person writes. When your 15 minutes are up, compare notes. Where are there similarities? Where are their differences? Why do those similarities and differences exist? What strategy did you each employ to take notes? How might you approach note-taking differently the next time you do it?

2. Let's simulate an ethnographic observation of a much more chaotic scene. While watching **Video 9.3**, write field jottings. (Write them by hand to simulate conditions in the actual field.) The video has some longer segments, but it is edited, which means you may find it hard to follow—don't worry about missing things. Try your best to capture as many details as you can. Then write up a page-long field note describing what you saw or heard. (In **Chapter 11: Qualitative Data Analysis**, we will analyze the data you collected, so hold on to your field note!)



One or more interactive elements has been excluded from this version of the text. You can view them online here: https://viva.pressbooks.pub/sociology-research-methods/?p=317#oembed-1

Video 9.3. January 6, 2021. *New Yorker* reporter Luke Mogelson was observing supporters of U.S. president Donald Trump as they protested outside the U.S. Capitol, challenging the results of the 2020 election. When some protesters smashed windows and forced their way into the building's entrances, Mogelson tagged along, using his phone to capture this footage of their actions within the building, including inside the U.S. Senate Chamber. You can use this video to practice ethnographic observation, as discussed in **Exercise 2**.

9.4. References

Armstrong, Elizabeth A., and Laura T. Hamilton. 2015. *Paying for the Party: How College Maintains Inequality*. Reissue ed. Cambridge, MA: Harvard University Press.

Best, Amy L. 2000. *Prom Night: Youth, Schools and Popular Culture*. New York: Routledge.

Best, Amy L. 2005. *Fast Cars, Cool Rides: The Accelerating World of Youth and Their Cars*. New York: NYU Press.

Best, Amy L., ed. 2007. *Representing Youth: Methodological Issues in Critical Youth Studies*. New York: NYU Press.

Best, Amy L. 2017. *Fast-Food Kids: French Fries, Lunch Lines, and Social Ties*. New York: NYU Press.

Boas, Franz. 1888. *The Central Eskimo*. Washington, DC: Government Printing Office.

Boelen, W. A. Marianne. 1992. "Street Corner Society: Cornerville Revisited." *Journal of Contemporary Ethnography* 21(1):11–51. doi: 10.1177/089124169202100100 2.

Burawoy, Michael. 1979. *Manufacturing Consent: Changes in the Labor Process under Monopoly Capitalism*. Chicago: University of Chicago Press.

Burawoy, Michael. 1998. "The Extended Case Method." *Sociological Theory* 16(1):4–33. doi: 10.1111/0735-2751.00040.

Burawoy, Michael. 2009. *The Extended Case Method: Four Countries, Four Decades, Four Great Transformations, and One Theoretical Tradition*. Berkeley: University of California Press.

Charmaz, Kathy. 2006. *Constructing Grounded Theory: A Practical Guide through Qualitative Analysis*. Thousand Oaks, CA: SAGE.

Chatman, Elfreda A. 1992. *The Information World of Retired Women:* Westport, CT: Praeger.

Chen, Katherine K. 2009. *Enabling Creative Chaos: The Organization Behind the Burning Man Event*. Chicago: University of Chicago Press.

Chen, Katherine K. 2012. "Charismatizing the Routine: Storytelling for Meaning and Agency in the Burning Man Organization." *Qualitative Sociology* 35(3):311–34. doi: 10.1007/s11133-012-9229-1.

Chen, Katherine K. 2015. "Using Extreme Cases to Understand Organizations." Pp. 33–44 in *Handbook of Qualitative Organizational Research: Innovative Pathways and Methods*, edited by K. D. Elsbach and R. M. Kramer. New York: Routledge.

Chen, Victor Tan. 2015. *Cut Loose: Jobless and Hopeless in an Unfair Economy*. Oakland: University of California Press.

Chen, Victor Tan, and Jesse Goldstein. 2024. "The Theatre of Entrepreneurship: Learning to Perform the Speculative Self in University Entrepreneurship Programs." *Social Problems* 71(1):203–19. doi: 10.1093/socpro/spac008.

Contreras, Randol. 2012. *The Stickup Kids: Race, Drugs, Violence, and the American Dream*. Berkeley: University of California Press.

Davis, Fred. 1973. "The Martian and the Convert: Ontological Polarities in Social Research." *Urban Life and Culture* 2(3):333–43. doi: 10.1177/089124167300200305.

Deegan, Mary Jo. 1988. *Jane Addams and the Men of the Chicago School, 1892–1918*. New Brunswick, NJ: Transaction Books.

Desmond, Matthew. 2016. *Evicted: Poverty and Profit in the American City*. New York: Broadway Books.

Didion, Joan. 2008. *Slouching towards Bethlehem: Essays*. New York: Farrar, Straus and Giroux.

Duneier, Mitchell. 1999. *Sidewalk*. New York: Farrar, Straus and Giroux.

Geertz, Clifford. 1973. *The Interpretation of Cultures: Selected Essays*. New York: Basic Books.

Glaser, Barney, and Anselm Strauss. 2017. *Discovery of Grounded Theory: Strategies for Qualitative Research*. New York: Routledge.

Glesne, Corrine, and Alan Peshkin. 1992. *Becoming Qualitative Researchers: An Introduction*. New York: Longman.

Gold, Raymond L. 1958. "Roles in Sociological Field Observations." *Social Forces* 36(3):217–23. doi: 10.2307/2573808.

Harrington, Brooke. 2020. *Capital without Borders: Wealth Managers and the One Percent*. Cambridge, MA: Harvard University Press.

Hochschild, Arlie Russell. 1989. *The Second Shift: Working Parents and the Revolution at Home*. New York: Viking Adult.

Horowitz, Mark, William Yaworsky, and Kenneth Kickham. 2019. "Anthropology's Science Wars: Insights from a New Survey." *Current Anthropology* 60(5):674–98. doi: 10.1086/705409.

Humphreys, Laud. 1970. *Tearoom Trade: Impersonal Sex in Public Places*. London: Duckworth.

Johanson, Graeme, and Kirsty Williamson, eds. 2013. *Research Methods: Information, Systems and Contexts*. Prahran, Australia: Tilde.

Junker, Buford Helmholz. 1960. *Field Work: An Introduction to the Social Sciences*. Chicago: University of Chicago Press.

Lareau, Annette. 2011. *Unequal Childhoods: Class, Race, and Family Life*. 2nd ed. Berkeley: University of California Press.

Liebow, Elliott. 1967. *Tally's Corner. A Study of Negro Streetcorner Men.* Boston: Little, Brown.

MacLeod, Jay. 2008. *Ain't No Makin' It: Aspirations and Attainment in a Low-Income Neighborhood*. 3rd ed. Boulder, CO: Routledge.

Malinowski, Bronislaw. 1922. *Argonauts of the Western Pacific: An Account of Native Enterprise and Adventure in the Archipelagoes of Melanesian New Guinea*. New York: Dutton.

McClennen, Sophia A. 2022. "'Taking Sides Is Not My Job': 'Retrograde' Filmmaker Matthew Heineman on Documenting a War's End." *Salon*. Retrieved January 23, 2024 (https://www.salon.com/2022/12/09/taking-sides-is-not-my-job-retrograde-filmmaker-matthew-heineman-on-documenting-a-wars-end).

Mitchell, Richard G., Jr. 1991. "Secrecy and Disclosure in Fieldwork." Pp. 97–108 in *Experiencing Fieldwork: An Inside View of Qualitative Research*, edited by W. B. Shaffir and R. A. Stebbins. Thousand Oaks, CA: SAGE.

Mitchell, Richard G., Jr. 2002. *Dancing at Armageddon: Survivalism and Chaos in Modern Times*. Chicago: University of Chicago Press.

Moskos, Peter. 2009. *Cop in the Hood: My Year Policing Baltimore's Eastern District*. Revised ed. Princeton, NJ: Princeton University Press.

Ocejo, Richard E. 2014. "Show the Animal: Constructing and Communicating New Elite Food Tastes at Upscale Butcher Shops." *Poetics* 47:106–21. doi: 10.1016/j.poetic.2014.10.006.

Ocejo, Richard E. 2017. *Masters of Craft: Old Jobs in the New Urban Economy*. Princeton, NJ: Princeton University Press.

Shaffir, William, Victor Marshall, and Jack Haas. 1980. "Competing Commitments: Unanticipated Problems of Field Research." *Qualitative Sociology* 2(3):56–71. doi: 10.1007/BF02390158.

Shankman, Paul. 2009. *The Trashing of Margaret Mead: Anatomy of an Anthropological Controversy*. Madison: University of Wisconsin Press.

Thorne, Barrie. 1993. *Gender Play: Girls and Boys in School*. New Brunswick, NJ: Rutgers University Press.

Vanderbilt University Writing Studio. 2012. "Tips for Writing Thick Descriptions for Ethnographies and Case Studies." *Vanderbilt University*. Retrieved January 23, 2024 (https://www.vanderbilt.edu/writing/resources/handouts/).

Viscelli, Steve. 2016. *The Big Rig: Trucking and the Decline of the American Dream*. Oakland: University of California Press.

Wallace, David Foster. 2009. *This Is Water: Some Thoughts, Delivered on a Significant Occasion, about Living a Compassionate Life*. New York: Little, Brown.

Wang, Tricia. 2016. "Why Big Data Needs Thick Data." *Medium*. Retrieved March 9, 2022 (https://medium.com/ethnography-matters/why-big-data-needs-thick-data-b4b3e75e3d7).

Whyte, William Foote. [1943] 1993. *Street Corner Society*. 4th ed. Chicago: University of Chicago Press.

10. IN-DEPTH INTERVIEWS



Interviews involve two or more people exchanging information through a series of questions and answers. Knowing how to conduct a good interview is a useful skill across a variety of professions. Market researchers use interviews to learn what potential or current customers think about their products. Police use interviews to extract information from witnesses and suspects. Journalists are masters of the art of interviewing, given how heavily they rely on this approach in crafting their stories. On some days, they might speak to dozens of people from a wide range of backgrounds—from celebrities and political leaders to random people on the street.



A reporter interviews veteran National Public Radio (NPR) journalist Linda Wertheimer after an event at Wellesley College in Massachusetts. Journalists are masters of the craft of in-depth interviewing, which they typically use to relate the views of public figures, convey expert information and analysis, and capture individual stories that characterize and humanize the issues they cover. *Graham Milldrum, via Flickr*

Sociologists use two types of interviews to gather data related to their research questions: *in-depth interviews* and *survey interviews*. Both allow us to learn what people feel about a particular issue, what their backgrounds and experiences are, and why they act or think in certain ways. However, these two approaches are very different in how they go about collecting data. *Broadly speaking, in-depth interviews prioritize depth, whereas survey interviews prioritize breadth.*

When sociologists employ **in-depth interviews**, they typically hope to hear what an **interviewee** thinks is important about a topic—and to hear it in their own words, and in detail. The questions for in-depth interviews are more likely to be **open-ended questions** like "How do you feel about marijuana legalization?" or "Why did you decide to attend college?" These questions can't be easily answered with a word or two. They encourage interviewees to think intently on a subject and describe their experiences, opinions, and attitudes *in depth* (hence the name of this approach). This is why in-depth interviews are used for qualitative research, and why they are also called *qualitative interviews*. Like

ethnographic observation, in-depth interviewing is intended to produce richly detailed information (as we've discussed before, **thick description**) that helps us understand complex phenomena.

By contrast, sociologists conduct **survey interviews** to produce quantitative data. This style of questioning reduces people's responses to simplified measures that can be tracked across groups and time periods. When using this approach, researchers tend to care less about fully understanding what one person in their sample thinks and more about seeing how people compare to each other in their responses. To help them discern patterns in people's diverse and sometimes complicated opinions, survey interviews pose **closed-ended questions**, asking people to choose from two or more set answers—much like multiple-choice questions or true/false questions do on a test. In this sense, the questions in survey interviews look more like those used in a courtroom cross-examination—not that they are hostile or meant to stump, but rather that they force the person being interviewed to answer briefly, without elaboration. By contrast, sociologists using in-depth interviews will pose their questions much like lawyers pose questions during a direct examination—asking the person on the stand to tell their story at length and in detail.

If a survey interview tends to have a formal feel to it—a question followed by a brief answer, one after another—a good qualitative interview can feel like a conversation, with few awkward pauses and both the interviewer and interviewee actively participating. That said, it is different from a normal conversation in several key ways. For one thing, the interviewer guides the conversation, asking questions that move the discussion toward areas relevant to the study at hand. That brings us to another key distinction between in-depth interviews and survey interviews. In an in-depth interview, you are not bound to your list of questions. You can follow up on new and unexpected topics that emerge during the conversation. Your questions can change as you learn more information. When you are conducting a survey, however, your questions must be *structured*: you want the same questions to be asked of your respondents, phrased in exactly the same way. Later in the textbook, we'll discuss why you might want to mix up your questions in a survey, but generally speaking, you need to take a consistent and standardized approach when conducting surveys. Even if a respondent's answer prompts you to think of a smart follow-up question, you cannot change course. Otherwise, the data you obtain from your survey will not be fully comparable across respondents, who might have been responding differently in part because they were asked more or less different questions.

In this chapter, we focus on in-depth interviews. (**Chapter 13: Surveys** covers survey interviews, in addition to the other, non-interview methods that researchers take to gather data for surveys.) We'll take a look at some of the strengths and weaknesses of in-depth interviews, talk about best practices for conducting them, and discuss issues of power and exploitation that come up when you work as intimately with research participants as you do when they are interviewees. Typically, in-depth interviews involve an in-person meeting between two people—the interviewer and interviewee—but interviews need not be limited to two people, nor must they occur in person. Qualitative interviews with more than one interviewee are known as **focus groups**. We will discuss this approach at the end of the chapter, as it presents unique advantages and challenges for researchers.

A note about terminology: sociologists use many terms to refer to the person the researcher talks to during interviews. As we've discussed in previous chapters, *research participant* is a catchall term for anyone involved in research—as the subject of an interview or an ethnographic observation, for instance—and you can rarely go wrong if you use that term. More specifically for interviews, however, you will hear terms like *informants*, *interlocutors*, *interviewees*, and *respondents*. The latter two terms are perhaps the most commonly used. We prefer "interviewee" because its meaning is obvious to nonacademic audiences. Using this terminology can also help us distinguish between people involved in (qualitative) interviews and those participating in (quantitative) surveys: "respondents" is consistently used to describe people answering questions for a survey ("the survey had 1,039 respondents"), whereas "interviewees" largely refers to participants in in-depth interviews.[1]

1. As usual, sociologists have differing opinions on this subject of what to call human subjects in interviews. For

Opening chapter image credit: *George Milton, via Pexels. Adapted by Bizhan Khodabandeh.*

quantitative studies, you usually will hear the term *respondent* when survey researchers refer to the people they interview or who fill out their forms; occasionally, *research participant* may be used as well. For qualitative studies, some sociologists prefer *informant*. However, this term is widely used in law enforcement circles to refer to a person with privileged information related to criminal activities—that is, a "criminal informant" (CI)—so it carries an implication that the individual has expert knowledge and is engaged in a long-term collaboration with the researcher, which may or may not be the case for a particular interviewee. The term *interlocutor*—academic jargon for a person involved in a conversation—is occasionally used, but like "respondent," it sounds rather clinical and formal. Note that when a study involves many interviewees, some may be distinguished as more important by being labeled as that study's *primary* or *key informants* or *focal participants* (for instance, if they are interviewed or observed more frequently than other individuals in the study).

10.1. When Should In-Depth Interviewing Be Used?

Learning Objectives

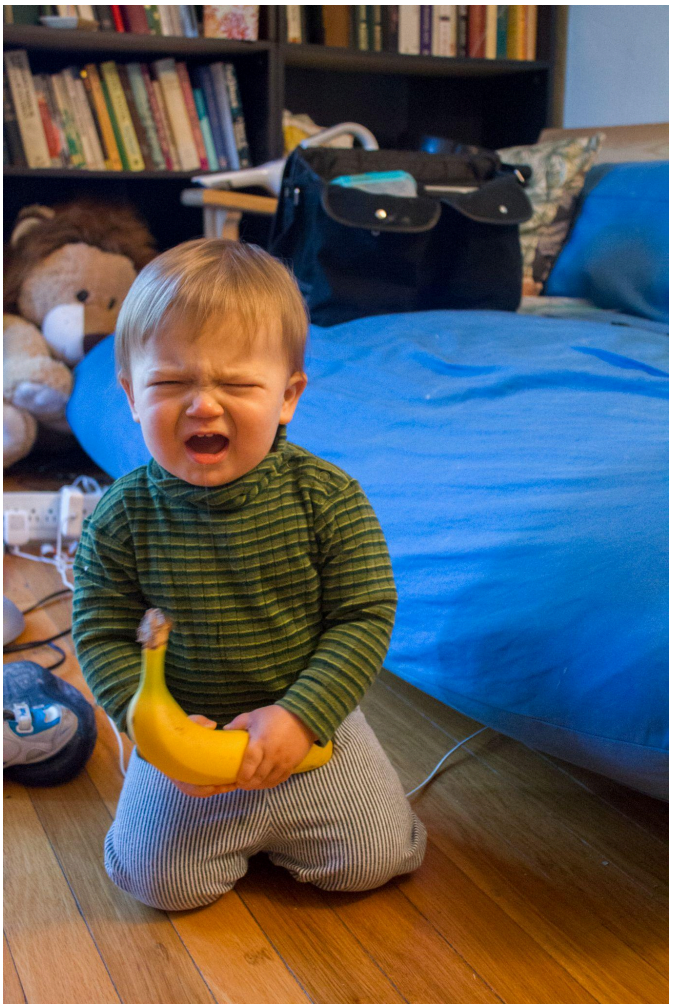
1. Identify when it is appropriate to employ in-depth interviews for data collection.
2. Discuss the drawbacks of in-depth interviewing as a research method.

In-depth interviewing is arguably the best research method to rely on if you want to explore a phenomenon in great detail. Not only are your interviewees given the opportunity to elaborate in a way that is not possible with other methods, but they also are able to share information in their own words and from their own perspectives. By contrast, survey research asks participants to fit their views into the limited response options provided by the researcher. Ethnographic observation can tell you how people act, and it can give you hints—through their expressions and mannerisms as well as their words—about what they are feeling in the moment, but in-depth interviewing provides easier access to people's subjective views, motivations, and emotions. Through what they relate during the interview, you as a researcher gain a window into how exactly they feel about a topic (or at least what they *say* about how they feel, as we'll discuss later).

Because qualitative interviews are designed to elicit detailed information, they are especially useful when a researcher's aim is to study the "how" of various phenomena. Put another way, they are adept at understanding **causal mechanisms**—the specific pathways by which one variable affects another variable (as we mentioned in **Chapter 3: The Role of Theory in Research**, qualitative research is particularly good at elucidating causal mechanisms). In part, this is because in-depth interviews allow interviewees to describe—at length, and in depth—the process by which a phenomenon occurs and the many dimensions it encompasses. For example, you could ask people about how exactly growing up in an affluent home with well-educated parents prepared them to pursue a professional career years later. Your interviewees could tell you stories of all the things they learned from a young age about the work world and what types of professions they could pursue, who or what they needed to know to get those jobs, and so on. If you are a skilled interviewer, your interviewees might provide many examples of such causal mechanisms and might explain each of them in great detail. Likewise, if you were interviewing people who decided not to have children, you could ask them how others in their lives have responded to that decision. As you can imagine, answering these "how" questions often requires a back-and-forth dialogue with interviewees. When they begin to tell you their story, inevitably new questions will come up—ones that may not have occurred to you when you did previous interviews, given that each person's story is unique. In-depth interviews allow you to pose these follow-up questions, but survey interviews do not.

More broadly, in-depth interviewing excels at gathering data about a process or a sequence of events. For instance, if you are interviewing people who decided not to have children, you might ask them how they came to that decision. They might recount one or more incidents or conversations that pushed them in the direction of not having kids, and you could capture that sequence in your in-depth interviewing. Because the process of making such an important decision can be complicated for many people—involving multiple overlapping and contradictory reasons—describing that process by responding to closed-ended questions on a survey wouldn't work particularly well. Through in-depth interviews, you can also gather data about events or experiences that are difficult or impossible for you to directly observe, including those that took place far back in the past. Indeed, one common strategy in in-depth interviewing is to ask about an interviewee's **life history**—the formative events in their lives (family life, education, work history, and so on)

that may have contributed to their orientation toward a particular issue or their experience of a particular phenomenon today.



In-depth interviews allow your research participants to recount processes and sequences of events. This can be helpful in studying potentially complex phenomena, such as how people make decisions. For instance, if you were trying to understand why a particular person doesn't want to have children, they could tell you on a survey that they cannot abide little kids, but you wouldn't be able to walk through their decision-making process with them—for instance, how multiple distressing incidents they experienced involving angry fruit-wielding toddlers led them to that view. *Quinn Dombrowski, via Flickr; banana for scale*

In-depth interviews are especially effective when you are dealing with **emergent phenomena**—that is, newly emerging topics or issues that researchers have not yet studied intensively. In an in-depth interview, your questions can plumb the various dimensions of that phenomenon, and you can adapt your questions as you learn more about it, asking follow-up questions as needed. As we described in **Chapter 4: Research Questions**, if you are conducting exploratory research on a brand-new development—say, a popular new app or a policy change—you might want to start off by conducting in-depth interviews with the people who use the app or those who have experienced the policy change. That would give you some basis for figuring out the issues and concerns you should really care about. You could then devise a new set of questions to be asked of a larger pool of people—either through in-depth interviews or surveys—based on what you discovered in your initial open-ended conversations.

An often-overlooked benefit of qualitative interviews is that researchers can make observations that go beyond what interviewees explicitly say. Their body language can give hints about their mood or their views on a topic they're discussing. For instance, a smirk or an awkward pause might suggest they have ambivalent feelings about the words coming out of their mouths (Pugh 2013, 2014). In addition to these **nonverbal cues**, the interviewee's appearance and overall demeanor can give you useful information about their social position and relationship to the topic you're studying. Even their choice of a particular time and location for an interview might say something meaningful, depending on the topic you're studying. Indeed, as we'll discuss later, if the site where you hold the interview has some significance to the interviewee—such as their home or workplace—taking detailed notes about that setting will allow you to garner valuable details about who they are and what they do or think.

As with any research method, the benefits of in-depth interviews come with drawbacks. First, there is the problem of veracity. Qualitative interviews rely on interviewees' ability to accurately and honestly recall specific details about their lives, circumstances, thoughts, opinions, and behaviors. Through interviews, you can learn what people think about a particular issue—but their stated views may be erroneous or deceptive. You can explain why people acted the way they did—but the reasons they give may not be correct.

Obviously, some people are willing to lie to an interviewer if they do not want to divulge certain information or they want

to present themselves in a better light (the latter being an example of **social desirability bias**, which we discussed in **Chapter 7: Measuring the Social World**). But interviewees can mislead researchers even when they do not intend to do so. If your questions have to do with something far in the past, or if you're asking about precise details, interviewees will often get things wrong. Eyewitness testimony is notoriously unreliable, given how memories get distorted over time to line up with people's views or assumptions. People may truly believe they do things for particular reasons, yet empirical evidence may suggest other considerations are more decisive—consider, for instance, the findings in social movements research that people get involved with activism largely because of their social ties to other people who are already part of those movements, rather than, say, their undying personal passion for a particular cause (McAdam 1990). And interviewees often need some sort of documentation to help them recall specific figures—for instance, the household income they reported on their taxes. For all these reasons, ethnographic observation and qualitative content analysis can be superior to in-depth interviewing if your goal is to get trustworthy data on what people actually do rather than what they say they do (Esterberg 2001). With firsthand observation, you'll actually see how they act. With a content analysis of video footage or written records, you'll hopefully gain a picture of their activities that they cannot so easily airbrush or hide.

Another drawback of qualitative interviewing is that it requires a lot of time and effort. In addition to drafting questions, recruiting interviewees, and conducting interviews, you usually want to write down notes about your observations and transcribe or at least summarize what was said in each interview. You then need to analyze the qualitative data you collected, which can be an especially painstaking process. Keep in mind you are also asking much more of your interviewees than you would be if you simply emailed them a questionnaire with closed-ended questions. Among other things, this high bar for participation means that fewer people will take you up on your offer to sit down for an in-depth interview.[1]

As we discussed in **Chapter 6: Sampling**, the intensity of in-depth interviewing means that you almost never can gather a representative sample of your population of interest. You will therefore need to look with skepticism at any findings you uncover from your nonprobability sample. For the most part, you will just want to conclude that any pattern you observed in your qualitative data *may* apply more broadly—that is, your results *suggest* this might be true—and leave it to other researchers to suss out whether the pattern holds up in other contexts (refer to **Chapter 4: Research Questions** regarding the limitations of the inductive approach that qualitative researchers typically employ, as well as the ways they can move beyond that approach of strictly theory generation). That said, rigorous in-depth interviewing can provide a compelling and credible picture of social life. And many qualitative researchers pair in-depth interviewing with other methods, such as ethnographic observation and content analysis, so that the data obtained through other research methods can be used to corroborate or challenge their interview data (and vice versa)—a process known as **triangulation**.

Conducting qualitative interviews not only is labor-intensive but also can be emotionally taxing. Hearing about the negative impacts that social problems have on people is difficult, especially when you are learning about them directly from the people affected. You might find that people get upset when you broach a sensitive topic, and they may even become hostile toward you. Researchers new to in-depth interviewing should consider what reactions they might

1. Because they demand so much more from both the researcher and research participant, studies that rely on in-depth interviewing and other qualitative methods tend to have major problems with *nonresponse bias* (as we discussed in **Chapter 6: Sampling**, this bias becomes a problem when the people who agree to be part of our study look meaningfully different from our target population, skewing our study's findings). This nonresponse bias can make it futile to obtain a truly representative sample, which is one reason that qualitative interviewers tend to favor a purposive strategy for selecting interviewees.

encounter in the course of asking their questions, and whether they are in the right state of mind to listen without judgment to what people have to say.

To summarize, in-depth interviews are especially useful when the following are true:

1. You wish to gather very detailed information.
2. You anticipate wanting to ask respondents follow-up questions based on their responses.
3. You plan to ask questions that require lengthy explanation.
4. You are studying a new, complex, or potentially confusing topic.
5. You are studying processes, such as how people make decisions.
6. You want to learn about events or experiences that are difficult for you to directly observe.

Key Takeaways

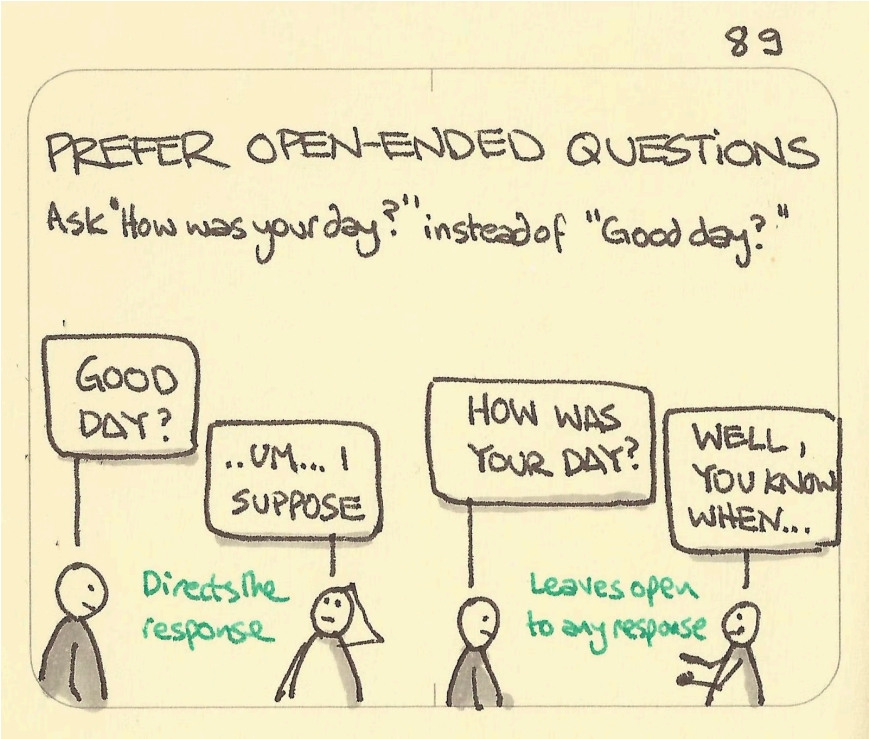
1. Researchers typically use in-depth interviews when they need to obtain highly detailed information, such as when they want to understand complex processes or emergent phenomena.
2. In-depth interviews may not provide the most trustworthy data—not just because interviewees may omit or misrepresent facts, but also because they simply may not know or recall the correct answers.
3. Qualitative interviews are an intensive research method to pursue in terms of the time, effort, and possible emotional strain involved.

10.2. Creating Interview Guides

Learning Objectives

1. Describe the difference between structured and semi-structured interviews and how in-depth interviewing uses its lack of structure to the researcher's advantage.
2. Discuss the forms that interview guides can take and how they should be organized.
3. Identify some rules of thumb for devising effective qualitative interview questions.

As we noted, survey interviews are **structured interviews**. They ask the exact same questions in a precise ordering that the researchers planned ahead of time. By contrast, most in-depth interviews are **semi-structured interviews**. The interviewer has a list of questions prepared in advance, but they can deviate from that list whenever they wish. That means the interviewer's questions may not be asked in exactly the same way or in exactly the same order to each and every interviewee. More rarely, in-depth interviews can be entirely **unstructured interviews**: no list of questions exists, and instead, the interviewer improvises questions as they see fit regarding a particular topic they want to hear about.



Unlike surveys, in-depth interviews prioritize *open-ended questions*—questions that do not include set answers to choose from. As a result, they demand much more attention from interviewees, who cannot simply pick one of the available answers when responding. *Jono Hey, via* Sketchplanations

For semi-structured interviews, sociologists develop an **interview guide**, a list of topics or questions that they can refer to during the interview (or memorize beforehand). It is called a "guide" because it is simply that: it is used to guide the interviewer, but it is not set in stone. Think of an interview guide as a to-do list—it contains all the tasks you hope to accomplish, but it probably won't be the end of the world if you don't check off every task or don't do them in the exact order you have written down. New tasks may come up that require you to adjust your plans, and you may not have the time that day to get to everything on the list.

When you are preparing your interview guide, you should outline all the issues that you believe are likely to be relevant during the course of your conversation with an interviewee. That said, because you will be asking interviewees to provide answers in their own words and raise any points they believe are important, each interview will flow differently. Although the opening question in an in-depth interview may be the same across all interviews, from that point on, what the interviewee says will shape how the conversation proceeds. This is what makes in-depth interviewing exciting—and rather challenging. It takes a skilled interviewer to be able to pose questions and probe more deeply into a topic as needed even while listening attentively to what an interviewee is saying at any given moment. The researcher needs to be able to pick up on cues about when to follow up, when to move on, and when to simply let the participant speak without guidance or interruption.

We've said that interview guides can list topics or questions. A topic-based interview guide will briefly outline the subjects the researcher would like to discuss, usually using just a word or phrase to describe each one. Those keywords will give the researcher some idea about how to generate questions on the spot. **Figure 10.1** provides an example of a topic-based interview guide for a study focused on the relationship between substance abuse and unemployment. As you can see, researchers can order their outline according to which topics should be addressed at the beginning of the conversation and which at the end, though they may very well move back and forth between topics and otherwise deviate from that ordering over the course of the interview.

Background
Born? Grow up? Neighborhood?
What was it like? Kind of people?
Lived with? Their work? Schooling?
What were they like? Stable?
Depressed? Mentally ill? Suicidal?
Drank? Used drugs/meds? Jailed?
Lose someone close to you?
Felt safe and secure? Why?
Not enough food, clothes?
Someone loved/felt you're special?
Abuse? Threats? Physical/sexual?
Swear at you or put you down?
Ever move? What like? Disruptive?
Ever homeless? What happened?
Relationship with family? Friends?
Grandparents? Their education/work?

Household
Where live? How long?
Neighborhood like? Safe?
Who live with? Children?
Who cares for kids?
Partner? How met?
Their work? Education?
Help you pay bills? Amount?
Previous marriage etc.? When?
Cars? License? Insurance?

Education
Degrees? Schools? Training?
Enjoyed school? Go back?
Friends? What circles/classes?
Kept in touch? Help? Benefits?
Military? Rank? Role? Benefits?

Employment
Working? Hustles, under table?
How got job? How long been there?
Job title? Hours? Pay? Enough?
PTO? Insurance? Pension? Union?
Responsibilities? Workplace like?
Like job? Enjoy work? Intense?
Management? Coworkers?
Opportunities? Why left/staying?
What age began working? [*History*]
Done gig work? Intern/volunteer?
Ever unemployed or without work?
Job search? Got help? Easy/hard?
Encountered scam/sketchy listings?
How made ends meet? Relaxing?
UI? Amount? Easy to get? Enough?
Career goals? Role models? Confident?

Health
Describe your state of health?
Chronic health issues? In family?
Injured at work? Workers comp?
Ever feel pain on regular basis?
1-10? Affect work? Relationships?
Prescribed pain meds or narcotics?
Why given? Dosage? Amounts?
Hard to stop? Concerned about?

Health insurance and disability
Insurance? Medicaid/ACA/Veterans?
Happy with? Copay/deductible?
Covers meds? Dental/vision?
Ever had to delay or not get care?
Afford treatment for SUD, therapy?
Received disability? When? Amount?
Last visited doctor? ER? For kids?

Substance history
Initial attitude about drugs?
Who in your life drank/used?
How old when first drank?
How obtained? Enjoy it?
Period when drank a lot?
How often? What kind?
When alone? To get thru day?
Withdrawals? How avoid?
Family/friends knew?
Affected relationships?
Affected ability to work?
How old when first used?
[*Ask same questions*]
Different from drinking?
Led to other drugs? Multiple?
Hard to get? Trust person?
Worry about OD? Narcan?

Recovery
Tried to get clean/sober? Relapse?
What was/is it like being clean?
12-step? MAT? IOP? Residential?
Would consider? Required?
Where? Got how far in program?
Describe? Like/dislike? Effective?

Emotional well-being
Describe your emotional health?
Happy person? Satisfied with life?
What causes stress? Excitement?
Sense of shame because SUD?
Self-blame? "Addict/loser/junkie"?
Problems sleeping? Hours/day?
Ever depressed? Anxious? Cause?
Considered suicide? Ever TDO'd?
Taking meds? Therapy? Effective?

Financial hardship
How difficult to pay your bills?
Ever unable to pay on time or in full?
Rent behind? Utilities cut? Repo?
Hard to cover $400 emergency?
Ever gone without necessities?
Cutting back on expenses? How?
Debt? Student loans? Credit cards?
Payday, title, tax lending? Pawning?
How many credit cards? How use?
Worried about finances/debt?
Ever declared bankruptcy?
At risk of losing home? Evicted?
Family/friends provide help? How?
Cash, groceries, gifts, place to stay?
How doing compared to parents?

Is there anything I missed that you think would be important or useful for me to know?

Is there anyone else you think I should talk to?

Legal history
Ever arrested or charged?
Drinking/using? Other?
ASAP? Post bail/bond? Help?
Incarcerated? How long?
Experiences with facility?
Inmates? Guards? Working?
Treatment? Other support?
Using during/after prison?
Affected you/sobriety? Family?
Fees from court, incarceration?
Amount? What for? Difficult?
Affected work? ID? Probation?

Government or charity help
SNAP? Food pantries?
Section 8? Rent? Shelter?
Pell grants? Veterans?
Healthcare? TANF? EITC?
WIC? Child/care credits?
Child support? Alimony?
Clothing? Charities?
Easy to get? Feel about help?

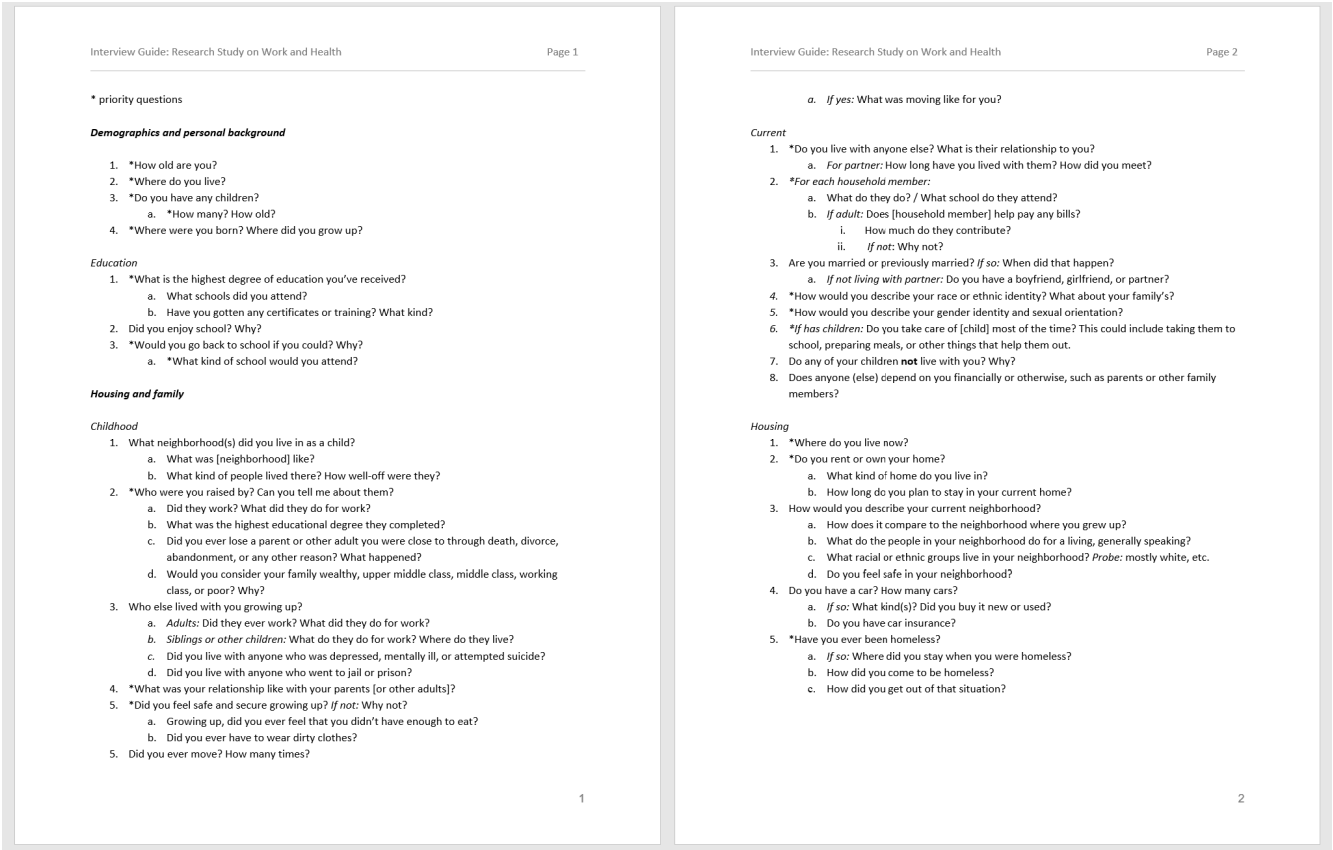
Social networks
Anyone to turn to re: problems?
Close to partner? Kids? Family?
Friends? Where live? Education/jobs?
Talk/see often? How keep in touch?
Online? What talk or do together?
How doing compared to parents?
Do for fun? Games? Last vacation?
Kids: School? After? Summer?
What does future have in store?
Involved in community? Go to church?
Campaigns? Do what? Describe faith?

Attitudes
Level of interest in politics? Why?
Which party do you support?
Conservative, liberal, moderate?
Voted in last election? Vote(d) for?
Where place yourself (10 at top)?
What is needed to be successful?
How is govt. handling economy?
Trade? Immigration? Inequality?
Incomes too unequal in country?

Respondent code: _____

Figure 10.1. Interview Guide with General Topics. This example interview guide (relating to a study drawing on in-depth interviews with economically precarious individuals who use opioids) contains only topics rather than specific questions. A topic-based interview guide assumes the interviewer will improvise the specific phrasing of their questions during the actual interview. It leads to a more natural conversation, but it also means that the interviewer needs to know their material well.

A question-based interview guide will frequently be organized by topics, too, but under each topic heading will be a number of detailed questions, written out more or less as the interviewer will say them aloud. The topics and questions will also be listed in the order that they are likely to proceed. **Figure 10.2** provides an example of a more detailed interview guide with specific questions posed.



Interview Guide: Research Study on Work and Health Page 1

* priority questions

Demographics and personal background

1. *How old are you?
2. *Where do you live?
3. *Do you have any children?
    a. *How many? How old?
4. *Where were you born? Where did you grow up?

Education

1. *What is the highest degree of education you've received?
    a. What schools did you attend?
    b. Have you gotten any certificates or training? What kind?
2. Did you enjoy school? Why?
3. *Would you go back to school if you could? Why?
    a. *What kind of school would you attend?

Housing and family

Childhood

1. What neighborhood(s) did you live in as a child?
    a. What was [neighborhood] like?
    b. What kind of people lived there? How well-off were they?
2. *Who were you raised by? Can you tell me about them?
    a. Did they work? What did they do for work?
    b. What was the highest educational degree they completed?
    c. Did you ever lose a parent or other adult you were close to through death, divorce, abandonment, or any other reason? What happened?
    d. Would you consider your family wealthy, upper middle class, middle class, working class, or poor? Why?
3. Who else lived with you growing up?
    a. *Adults:* Did they ever work? What did they do for work?
    b. *Siblings or other children:* What do they do for work? Where do they live?
    c. Did you live with anyone who was depressed, mentally ill, or attempted suicide?
    d. Did you live with anyone who went to jail or prison?
4. *What was your relationship like with your parents [or other adults]?
5. *Did you feel safe and secure growing up? *If not:* Why not?
    a. Growing up, did you ever feel that you didn't have enough to eat?
    b. Did you ever have to wear dirty clothes?
5. Did you ever move? How many times?

1

Interview Guide: Research Study on Work and Health Page 2

    a. *If yes:* What was moving like for you?

Current

1. *Do you live with anyone else? What is their relationship to you?
    a. *For partner:* How long have you lived with them? How did you meet?
2. **For each household member:*
    a. What do they do? / What school do they attend?
    b. *If adult:* Does [household member] help pay any bills?
        i. How much do they contribute?
        ii. *If not*: Why not?
3. Are you married or previously married? *If so:* When did that happen?
    a. *If not living with partner:* Do you have a boyfriend, girlfriend, or partner?
4. *How would you describe your race or ethnic identity? What about your family's?
5. *How would you describe your gender identity and sexual orientation?
6. **If has children:* Do you take care of [child] most of the time? This could include taking them to school, preparing meals, or other things that help them out.
7. Do any of your children **not** live with you? Why?
8. Does anyone (else) depend on you financially or otherwise, such as parents or other family members?

Housing

1. *Where do you live now?
2. *Do you rent or own your home?
    a. What kind of home do you live in?
    b. How long do you plan to stay in your current home?
3. How would you describe your current neighborhood?
    a. How does it compare to the neighborhood where you grew up?
    b. What do the people in your neighborhood do for a living, generally speaking?
    c. What racial or ethnic groups live in your neighborhood? *Probe:* mostly white, etc.
    d. Do you feel safe in your neighborhood?
4. Do you have a car? How many cars?
    a. *If so:* What kind(s)? Did you buy it new or used?
    b. Do you have car insurance?
5. *Have you ever been homeless?
    a. *If so:* Where did you stay when you were homeless?
    b. How did you come to be homeless?
    c. How did you get out of that situation?

2

Figure 10.2. Interview Guide with Topics and Questions. In this example interview guide for the same study, questions are written out as the interviewer would pose them in the actual interview. (If you prefer, you can also just summarize the questions in abbreviated bullet points.) Note that although this interview guide includes questions, it also is organized by topics, which helps the interviewer move the conversation from topic to topic with a natural flow. During the actual interview, however, the researcher can ask questions out of sequence or generate new questions on the fly. Also note that this interview guide includes planned probes—questions asking for more information—that follow the main questions; while you can and should improvise probes during the actual interview, it is a good idea to plan some in advance.

You may have noticed that some of the questions in the example interview guide in **Figure 10.2** have additional questions listed after or below them. These follow-up questions are known as **probes**. Remember that one of the benefits of qualitative interviews is that you can ask participants for more information. Especially if you ask any yes/no questions or similar closed-ended questions, be sure to ask a follow-up question or two afterward to get your interviewee to expand on their previous answer. Probes are routinely improvised during the course of an interview, and we discuss simple

strategies for generating probes on the spot later in the chapter. That said, it is worth your time to come up with helpful probes in advance of an interview (perhaps listing them, as in the example, as subquestions). That way, you have some probes already on tap and won't get stumped about how to follow up if your interviewee suddenly falls silent.

Whether you use an interview guide with written-out questions or just topics is up to you. In deciding between these options, consider your style, experience, and comfort level as an interviewer as well as your familiarity with the issues being discussed. If you are new to interviewing or the topic at hand, writing out the questions in advance may work better. Another option is to write up a detailed guide that you use to help you prepare for your interviewing, but using just an outline of topics for your actual interviews. Drafting specific questions for the interview guide will get you in the mindset of thinking creatively about the questions you can ask and will also ensure that you don't leave out anything important. You can use the question-based guide to practice interviewing and even rely on it for the first few interviews you conduct. At some point, however, we recommend transitioning to just using a topic-based guide. Relying on an outline will encourage you to listen to what your interviewee is telling you rather than just robotically moving to the next question on your list. An overly detailed interview guide may be difficult to navigate during an interview and could give interviewees the impression that you are more interested in your questions than in their answers.

When beginning to construct an interview guide, brainstorming is usually the first step. There are no rules at the brainstorming stage—simply list all the topics and questions that come to mind when you think about your research question. Once you've drafted a pretty good list, you can begin to tighten up the wording of your questions and flesh out important parts of your outline. Remember to come up with thematic headings and group together "like" questions and topics under those categories, which will make it easier to see where new questions should go as well as which redundant questions can be dropped.

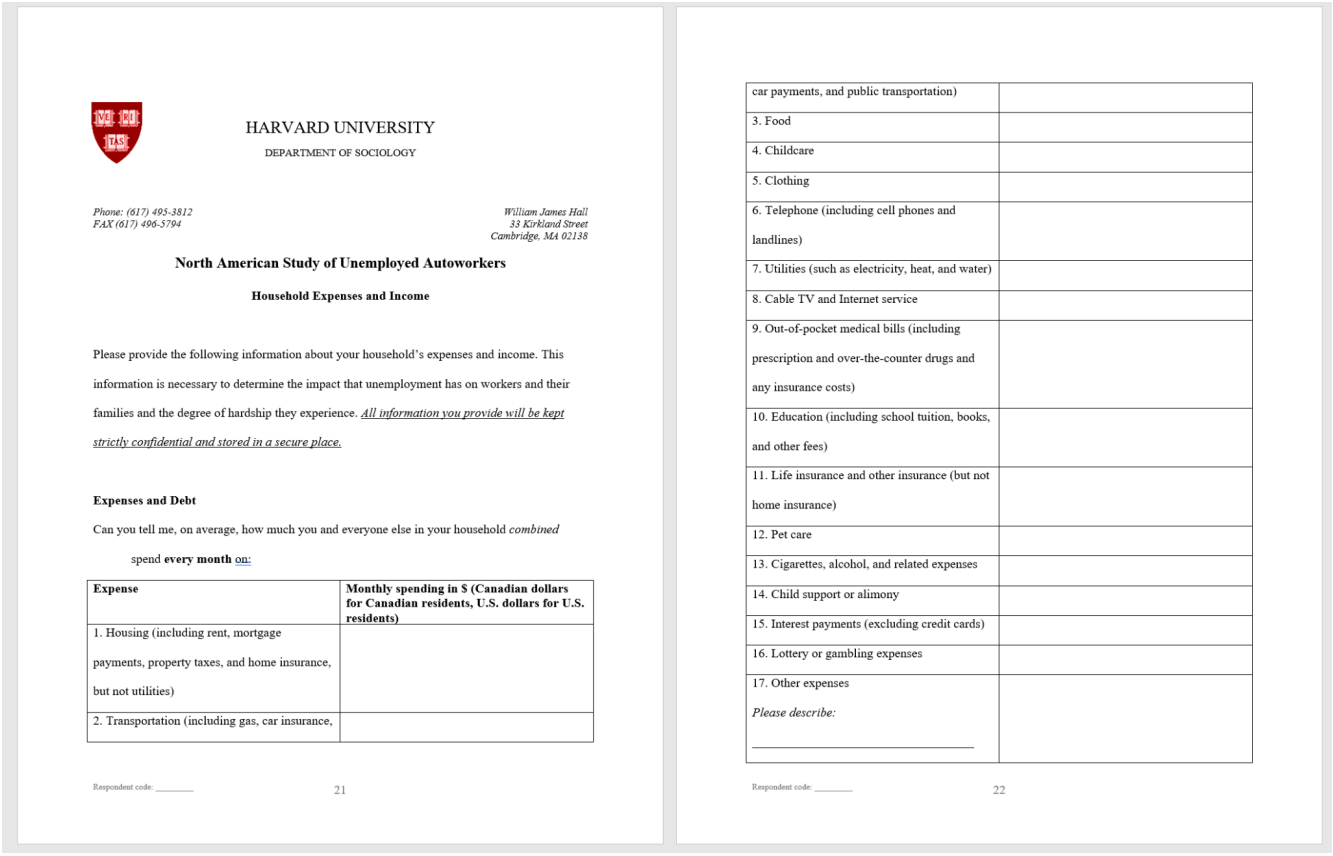
Eventually, you should figure out a logical ordering to your topics and questions. That way, the conversation flows smoothly from one theme to another, and both the interviewer and interviewee have some sense of what you have already covered and where you are headed. The ordering you choose for your interview guide may in part be chronological—for instance, you could start by talking about an interviewee's relevant personal background, and then move to what is going on now. Other than that, though, you just want a sensible sequence to the conversation, without jarring transitions—for instance, moving directly from asking about their relationship to their parents to asking about their sex life. Indeed, if you need to pose sensitive or potentially controversial questions, try your best to find a natural segue into these topics, and do not raise them at the very beginning of your interview. You need to give participants the opportunity to warm up to the interview and feel comfortable talking with you.

What sorts of questions should go into your interview guide? First, as we've emphasized, qualitative interviews focus on *open-ended questions*, questions that a researcher poses but does not provide answer options for. Open-ended questions are more demanding than closed-ended questions are. They require participants to put their thoughts and feelings into their own words. They are the sorts of questions that play to the strength of in-depth interviewing—specifically, its ability to generate rich qualitative data and obtain a nuanced understanding of the issue at hand. To this end, try to phrase your questions in a way that invites longer responses from your interviewees. Minimize the number of questions that can be answered with a simple yes or no; if you do use such questions, try to pair them with probes that are open-ended.

That said, you do want to include questions that ask about demographic and other important background characteristics—for instance, your interviewee's age, gender, racial and ethnic background, marital status, and political affiliation. These can become the basis for generating interesting theories. For instance, you may find that people from more privileged class backgrounds tend to have stronger opinions about a particular issue, or men are less willing to broach a sensitive topic.

Some qualitative interviewers create a brief survey form that interviewees can fill out with their answers to these background questions. That way, the interviewee can take care of these questions before or after the interview, and

the two of them do not have to go over them during their conversation. This approach of using a **supplemental questionnaire** may be especially helpful if you are asking interviewees to divulge private details or recall specific details. For instance, interviewees may be more willing to disclose information about their sexual history, such as the number of sexual partners they have had, on a survey form rather than saying it aloud. Likewise, they may not remember precise figures, such as household income or expenses, on the spot, but they could provide accurate information at a later opportunity when they can refer to their records (be aware, though, that you may need to pester them later about turning in the form!). **Figure 10.3** illustrates what such a supplementary questionnaire looks like.



HARVARD UNIVERSITY
DEPARTMENT OF SOCIOLOGY

Phone: (617) 495-3812
FAX (617) 496-5794

William James Hall
33 Kirkland Street
Cambridge, MA 02138

North American Study of Unemployed Autoworkers

Household Expenses and Income

Please provide the following information about your household's expenses and income. This information is necessary to determine the impact that unemployment has on workers and their families and the degree of hardship they experience. *All information you provide will be kept strictly confidential and stored in a secure place.*

Expenses and Debt

Can you tell me, on average, how much you and everyone else in your household *combined* spend **every month** on:

Expense	**Monthly spending in $ (Canadian dollars for Canadian residents, U.S. dollars for U.S. residents)**
1. Housing (including rent, mortgage payments, property taxes, and home insurance, but not utilities)	
2. Transportation (including gas, car insurance, car payments, and public transportation)	
3. Food	
4. Childcare	
5. Clothing	
6. Telephone (including cell phones and landlines)	
7. Utilities (such as electricity, heat, and water)	
8. Cable TV and Internet service	
9. Out-of-pocket medical bills (including prescription and over-the-counter drugs and any insurance costs)	
10. Education (including school tuition, books, and other fees)	
11. Life insurance and other insurance (but not home insurance)	
12. Pet care	
13. Cigarettes, alcohol, and related expenses	
14. Child support or alimony	
15. Interest payments (excluding credit cards)	
16. Lottery or gambling expenses	
17. Other expenses *Please describe:* _____	

Respondent code: _____ 21

Respondent code: _____ 22

Figure 10.3. Supplemental Questionnaire. Rather than spending time in your interview asking your interviewee about basic demographic information, you can use a survey form to capture those details, which the interviewee can fill out before or after the two of you talk. You can also use supplemental questionnaires to ask sensitive questions (which interviewees might not want to answer out loud) or to obtain precise figures (which interviewees may not know off the top of their head but can find out later). This example questionnaire was for Victor's study of long-term unemployment (Chen 2015) and asked interviewees for specific details about their household expenses and income.

Devising Effective Questions for In-Depth Interviews

You want to pose questions in an in-depth interview that answer your research question and cover all the relevant aspects of the phenomenon you're studying. When you are brainstorming, think of questions that touch upon the various concepts and theoretical relationships that your research question mentions or implies. Consult the scholarly literature to find out what questions other interviewers have asked in studies of similar topics. To make sure your interview guide is getting at the full range of issues your study seeks to illuminate, you may even want to create an

annotated version of your guide for planning purposes that explicitly states what the point of each question is—what theories or concepts it addresses, what variation in a variable it is measuring, and so on.

We have three recommendations for generating theoretically informed questions that are likely to produce useful data. The first is to simplify your language. Think about the theories and concepts you are interested in and how you might break down these abstract ideas into scenarios or statements that are clear, concrete, and vivid to your interviewee. Avoid jargon like the plague. Academic lingo will utterly confuse many of your interviewees—and if it doesn't, it will likely signal to them that you have a particular take on the topic, encouraging them to answer accordingly.

Another suggestion is to think backwards about each question you plan to pose. How do you imagine people will answer this question? How would you answer it? If the answers you expect to receive won't produce the sorts of data that are helpful in answering your research question, then you should rephrase it or ask additional questions that will get you the data you need. For instance, if you are interested in what sorts of conditions lead to sexual harassment in a workplace, you may find yourself tempted to just ask that question directly of people you're interviewing who have themselves experienced sexual harassment. The question might generate some interesting responses, but it gives you only their take on the issue. You also should ask separately about the wide range of conditions that might have possibly contributed to the sexual harassment they experienced, to see which conditions seem decisive across the experiences of all the interviewees you speak to—whether or not they personally brought them up as being important.

You also want a useful degree of variation in how people answer your questions. Ideally, the interview conversation will cover all the relevant aspects of the phenomenon you're studying, which means your questions as a whole will provoke a wide range of answers. That variation will allow you to see patterns in how people answer questions. As we noted earlier, this is where your demographic and other more closed-ended questions become useful: you may find that people with different backgrounds or beliefs answer a key question in different ways, which becomes the basis for generating a theory. Therefore, when you think backwards about the answers you expect from a particular question, consider whether it gives you the variation you want. If you imagine everyone will answer one question in the same way (e.g., "Do you like puppies?"), is their unanimity of opinion important for you to flag? If not, maybe you want to ask a question that is less obviously answered. At the very least, you will want to ask probes to get at the reasoning behind their answer ("So why do you like puppies?") or unearth other considerations where you can observe real variation.

Third, avoid pushing your interviewees to answer in certain ways. Don't phrase your questions in a leading way, which will introduce bias into your research by making it more likely that you get the results you're looking for. For example, a question like "Don't you think people who don't want kids are selfish?" clearly implies that the appropriate answer is "Yes." Asking, "Do you think people who don't want kids are selfish?" is better, although you are still putting the idea in the interviewee's head that not choosing to have children is a selfish act—which may or may not be a major problem, depending on the focus of your study. A more neutral question would be something like, "What comes to mind for you when you hear someone doesn't want kids?" That allows the interviewee to bring up selfishness if they really believe it is a problem, or not mention it at all—a sign it isn't an important issue to them. (You can always ask a follow-up question about selfishness if the person didn't mention it and you want to get their take on that specific topic.) Note that even bringing up a topic earlier in your interview can prime people to think in those terms later on. For instance, if you talk about a person's attitudes about gender inequality earlier in the interview, and then ask people about issues in their workplace, they might be more likely to make a connection between those issues and gender dynamics at work.

Whenever you ask about behaviors that are stigmatized—anything from using illicit drugs to watching trash TV—people will naturally be inclined not to answer candidly because of social desirability bias. As we've discussed, this is an inherent weakness of in-depth interviewing as a method, but you can try to minimize it by asking questions in ways that make it more likely that an interviewee will actually offer the answer that doesn't make them look so favorably. For instance, when asking about his interviewees' mental health for a study he did on long-term unemployment, Victor would preface his question with a statement like this (see the corresponding page in his interview guide in **Figure 10.4**):

> Being unemployed is often a tough experience for people to go through. How is your mental or emotional health these days? Is it poor, fair, good, or excellent?

The introductory statement, "Being unemployed is often a tough experience for people to go through," normalizes what might be seen as embarrassing or stigmatized—having poor mental health. Knowing that it's common to feel anxious or depressed, the interviewee will hopefully be more open to divulging any mental health problems if they do indeed feel that way, Also note that there is a follow-up question—only asked after the primary question about mental and emotional health—to characterize their mental health as "poor, fair, good, or excellent." The idea is that they may brush aside the primary question with a few words, not providing much information about how good or poor their mental health actually is. It's important to get them to be more specific. For instance, you could pose a probe after this closed-ended question, perhaps to ask the interviewee why they chose that response.

When prefacing your question with a statement, you might want to use an "either/or" structure that validates multiple possible responses. For instance, to get at a person's feelings about the shame of not having a job, Victor would pose the following question:

> Some people say they feel a sense of shame because they're unemployed. Other people say, no, it's not my fault so I have nothing to be ashamed of. What's your view?

By implying that the interviewee is not alone in thinking one way or the other, the question makes them more comfortable about providing a potentially stigmatizing answer.

V.A. Subjective well-being

1. Being unemployed is often a tough experience for people to go through. How is your mental or emotional health these days? Is it poor, fair, good, or excellent?

---

   a. Do you have any problems sleeping at night? _____

   b. Are you ever depressed or anxious? Do you have any thoughts of suicide?

---

2. Are you receiving any kind of counseling? _____
3. Have you been able to talk to anyone (else) about what you've been going through?

---

4. All things considered, how satisfied are you with your life as a whole these days, on a scale of 1 to 10 (with 1 being dissatisfied, and 10 being completely satisfied)? _____

   a. Why do you feel that way?

---

5. Do you think you are less satisfied with your life now than when you were working at [the auto company]? _____
6. Some people say they feel a sense of shame because they're unemployed. Others say, no, it's not my fault so I have nothing to be ashamed of. What's your view?

Figure 10.4. Asking Sensitive Questions. When dealing with sensitive topics where an interviewee might be reluctant to disclose how they really feel, you might consider adding a sentence or two before the actual question. In the statement, you would indicate that one or more answers are perfectly acceptable as responses, therefore reducing the risk that the interviewee will answer in a way that saves face (i.e., social desirability bias). On this page from an in-depth interview guide for Victor's study of long-term unemployment (Chen 2015), you can see two examples of this approach highlighted in the green and red boxes.

To wrap up our advice on posing interview questions, we suggest that you make it a habit to end your interviews with the following two questions:

1. Is there anything I missed that you think would be important or useful for me to know?
2. Is there anyone else you think I should talk to?

You can reword these questions as you see fit, but you will probably find one or both of them helpful, regardless of your specific study. The point of the first question is to ensure that your take on the phenomenon you're discussing hasn't left out important aspects or details that your interviewee might not have brought up explicitly during the interview—perhaps because they wanted to defer to your expert perspective on the topic. The second question helps you generate leads for more people to interview, which is typically something you want to do in an interview-based study (even if you do not plan to rely heavily on snowball sampling, as we will discuss later).

After you draft your interview guide, look it over. Cut any questions and topics that seem redundant or unnecessary. Consolidate similar questions. Convert questions that ask for specific or narrow details into probes or subquestions listed underneath questions that cover that topic more broadly. All of these strategies will tighten up your interview guide, making it more manageable to employ and making your questions easier to remember and elaborate on during the interview. You might also want to flag certain questions as priority questions—perhaps with an asterisk in the interview guide—so that if you are pressed for time, you know which questions you definitely need to ask.

When your interview guide is ready, get some feedback on it from your peers and other researchers. Chances are they'll catch a few things you hadn't noticed. Practicing the interview in advance is helpful, too. Ideally, you'll find a friend or two willing to participate in a couple of trial runs with you. Better yet, you may think of a friend or two who are similar in at least some ways to the people in your target population. Your mock interviewees may suggest specific ways you can improve your interview guide or your interviewing style, but even if they don't, you'll find that just practicing the interview with a real person on the other end of the conversation makes a huge difference. You'll get familiar with what questions you should be asking next at every point in the interview, and you'll start thinking up good probes, which will mean you're less likely to be speechless after an interviewee has just said something. Going back to our earlier recommendation regarding thinking backwards about questions, we advise that you pay particular attention to how your mock interviewees answered any interview questions closely related to your overarching research question. What sorts of data were you hoping to get from those questions, and what sorts of data did they actually give you? If there's a mismatch, consider whether another line of questioning would be more effective.

Key Takeaways

1. In-depth interviews are usually semi-structured, meaning that the researcher has topics and questions in mind to ask but their questions are open-ended and flow according to how the participant responds to each. This allows the researcher to cover a topic in great depth and exhaustive detail.
2. Interview guides can vary in format but should contain some sort of outline of the topics you hope to cover during the course of an interview, hopefully with a logical ordering of them in mind.
3. For in-depth interviews, prioritize questions that require interviewees to talk at length and in detail, and think up possible probes that will get them to elaborate on what they say.
4. To ensure your questions are effective in getting you the data you need, improve them by simplifying your language, thinking backwards about the usefulness of the answers you expect to get, and avoiding question wording that leads your interviewees to answer in certain ways.

10.3. Preparing for In-Depth Interviews

Learning Objectives

1. Describe the various approaches that sociologists can take to recruit a sample for in-depth interviewing.
2. Identify some of the key considerations to keep in mind when choosing a location for an interview.
3. Discuss the different types of data and insights that sociologists can generate by recording conversations and writing field notes.

Now that you have an interview guide, you are ready to start scheduling interviews. In this section we'll discuss some of the practical issues you'll need to attend to before you do your first interview: interviewee recruitment and compensation, interview locations, and interview recording and notetaking.

How to Recruit Interviewees

Before you can begin any interviews, you need to determine an effective sampling strategy for recruiting interviewees. We already discussed the purposive sampling approaches you will want to use for any qualitative study in **Chapter 6: Sampling**. We want to point out, however, a few considerations distinct to in-depth interviewing. First, recruiting interviewees usually takes two general approaches. The first is using **network-based referrals**. This might entail adopting the **snowball sampling** strategy we highlighted earlier in the textbook, which involves asking individuals you've interviewed to suggest other potential interviewees—in effect, utilizing existing social networks to recruit more people for your study. Another network-based approach to finding interview respondents is through *organizational referrals*. You can talk to people who work at a particular organization—say, a nonprofit organization or government agency—and then get them to refer you to members of the population that their organization serves.

Note that this is easier said than done: many organizations have legal or professional rules preventing them from sharing personal information. That said, given the degree of trust you need to establish to get anyone to talk to you—especially those from more marginalized communities—you should use whatever contacts you have available to you, and sometimes the staff at community organizations are the sorts of gatekeepers you need to win over to garner a sizable sample of interviewees. If you take this approach, one strategy is to first interview the leadership at these organizations. These *expert interviewees* can provide important context for your study, and after they've sat down for an interview with you and have learned something about your study and your intentions, they might be more inclined to help you recruit interviewees from your actual population of interest. Less formal settings where your target population hangs out—anywhere from churches to support groups to sports clubs—may also have leaders or members you can ask for leads about potential interviewees, and those individuals may be less bound by privacy requirements.

The second general approach to recruiting interviewees is through **venue-based recruitment**—postings about your study in more or less public spaces. These announcements can take online and offline forms. You can post flyers asking for study participants in settings such as coffee shops, libraries, and grocery stories. You can also spread the word on social media platforms, discussion boards, and listings sites. If you can get the permission of an organization that works with your target population, you may even be able to disseminate your announcement through that organization's mailing list or post it in their offices. Regardless of where you publicize your study, you should draft an announcement

stating the **selection criteria** to participate in your study (i.e., that you are looking for college students, people who work in the retail industry, people between the ages of 18 and 35, and so on). In addition to briefly describing your study and any compensation you offer, the announcement needs to include one or more ways to reach you—an email address, phone number, and possibly a link to a website where you supply more details about your study. (To guard your privacy, you can create email addresses, voicemail numbers, and even websites through Google's various apps and similar free-to-use products.) You may also find it helpful to design an eye-catching and professional-looking graphic that summarizes your announcement, which could take the form of an image you share online or a physical flyer you print out. See **Figure 10.5** for an example of what a simple recruitment flyer might look like.

You are more likely to get people to agree to be interviewed if you have funds to pay them. The amounts that sociologists pay vary widely, depending on how much financial support their study has, but anywhere from $20 to $50 per interview (or hour of recorded interview) is typical. Unfortunately, getting funding to pay interviewees isn't feasible for many researchers, which makes the job of recruiting a substantial sample harder but not impossible. With enough persuasion and persistence, you'll find people willing to sit down with you for an interview—after all, journalists get people to talk to them all the time without payment.[1]

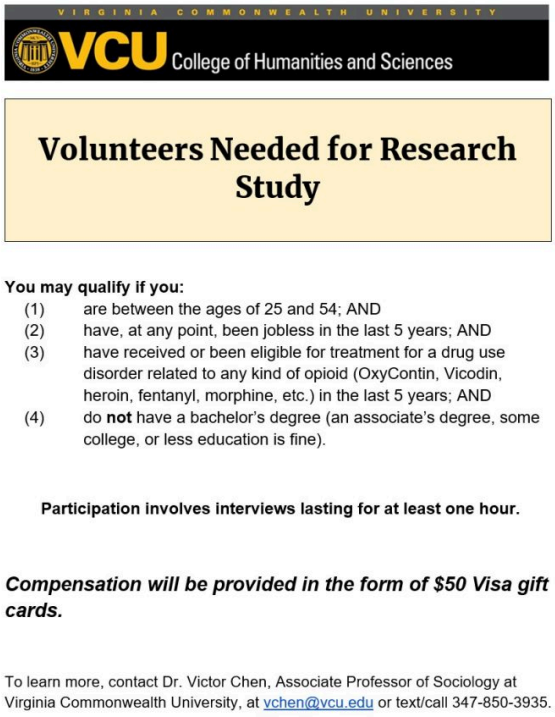


Figure 10.5. Recruitment Flyer. A recruitment flyer for a study on the relationship between drug use and unemployment. When drafting such an announcement, you should indicate the relevant *inclusion criteria*—what sort of background a volunteer needs to have in order to be included in the study—as well as the *exclusion criteria*—what personal characteristics rule out someone as a research participant (given the study's particular focus). You need to include your contact information so that volunteers can easily reach you, though you might set up an email address or phone number (say, through Google Voice) specifically for research purposes.

1. It's worth making note of how differently journalists and sociologists look upon the practice of compensating interviewees. Based on their profession's ethical norms, journalists largely oppose paying for interviews, fearing

Where to Conduct Your Interviews

Once you have recruited interviewees, you will need to decide where to conduct your interviews. In general, pick a location where your interviewee feels comfortable and is therefore more inclined to open up and speak truthfully. Oftentimes, the best location is someone's home. That way, they are not worried about other people they don't know or trust overhearing anything sensitive they say. However, some interviewees will be understandably wary about having strangers in their home. Depending on the topic, the presence of family members or housemates may also make them reluctant to talk openly.

Another possibility is interviewing people where they work—or, for that matter, any place related to their work. If you're interviewing people about their jobs, you can get a better sense of their workplace environment and its interpersonal dynamics by conducting the interview there. (In that case, make sure to observe the person's workspace, if applicable, to get a sense of their personal relationship to their work.) On the other hand, an interviewee may be reluctant to speak candidly about their work lives—such as any issues they have with management—if they might be in earshot of anyone else. Furthermore, their superiors might not take kindly to conducting interviews on the premises.

Locations that are not people's homes or workplaces—so-called *third places*—offer a wide range of additional possibilities for interview sites: anything from street corners and parks to coffee shops and libraries to churches and agencies. If you found your interviewees with the help of an organization, that organization may be willing to let you do interviews in one of their conference rooms or another space they control. This may give you further opportunities to observe the organization itself and your target population's interactions with their staff, fellow clients, and others.

potential conflicts of interest as well as the practical problems it introduces, such as encouraging bidding wars over highly sought-after interviewees. By contrast, sociologists typically seek to pay interviewees whenever possible. Based on another set of ethical priorities, they find it important to compensate people for their time and effort, which not only incentivizes participation but also helps make the exchange between interviewer and interviewee a bit more reciprocal and equitable (a topic we'll return to later).



An interview with a worker at a pine nut processing factory in Hongcheon, South Korea. If it is feasible to do so and your interviewees are comfortable with the idea, conducting interviews at their home or workplace allows you to make additional observations about potentially relevant topics like their relationships with their coworkers or their personal history and family background. *Ramon Carrillo (APFCN), via Flickr*

While you may gain something from observations you make at the interview site, weigh those advantages against your ability to have a thoughtful conversation in that space. (After all, you can set aside time to do a separate observation of any relevant location, perhaps with the interviewee accompanying you.) Make sure that whatever site you choose has few distractions. For example, some coffee shops and restaurants are so loud that recording the interview can be a challenge. The screech of espresso machines may drown out your voices. Servers may routinely interrupt your conversation. Likewise, if you conduct interviews with parents in their homes, they may out of necessity spend more time attending to young children during an interview than responding to your questions (of course, depending on the topic of your research, the opportunity to observe such interactions could be invaluable). As an interviewer, you may want to suggest a few possible locations, and note the goal of avoiding distractions, when you ask your interviewees to choose a location.

You can also conduct interviews over the phone or a video chat application. We generally don't recommend phone calls, because you gather so much more information about the person's expressions and mannerisms when you can see them talking. You may lose that vital data about their mood and outlook when you just have audio to go by. That said, video chat applications have become so advanced and ubiquitous that they are a good substitute for in-person interviewing.

That brings us to a final point about location: always prioritize your own well-being as an interviewer. You may need to do your interviews over video chat, rather than in person, for health or safety reasons. Ultimately, your respondent's comfort and convenience has to be balanced against a particular location's accessibility to you as well as how safe and comfortable you will feel there. For example, there's no reason you should feel pressured to conduct an interview at the late-night meeting place your interviewee prefers, much less in the inner sanctum of their secret society. Not only might you fear for your safety, you might be too distracted to conduct a good interview.

Interview Recordings and Notes

Before you embark on your first interview, you should decide whether you plan to record the conversation, and how you will go about taking notes. Even the most diligent sociologist cannot write down everything that is seen or heard during an interview. It is also difficult to be truly present and observant if you are writing down everything your interviewee is saying. For this reason, you probably want to follow the standard sociological practice of audio-recording your in-depth interviews. Of course, not all interviewees will feel comfortable being recorded, and sometimes even the interviewer may feel that the subject is so sensitive that recording would be inappropriate. Keep in mind, too, that transcribing the audio recordings of your interviews on your own may take considerable time. (In **Chapter 11: Qualitative Data Analysis**, we go over the transcription process as well as the automated transcription tools that are now available to shorten this process.) Sociologists often buy digital recorders for their interviewing, but today's smartphones have voice memo apps with very good audio quality. Video recordings are also an option, and they are especially helpful for interviews with multiple interviewees—and especially easy to create using video chat software if you're doing a remote interview.

Regardless of whether you record the interview, we do recommend that you write **field jottings**—research notes—during your interviews (and before and afterward as well, if there is anything interesting to observe). As we described in **Chapter 9: Ethnography**, you should then convert your jottings into a formal **field note**. Field notes help researchers document what they observe, and in so doing, they form the first draft of data analysis. Field notes for in-depth interviews may contain many things—observations of body language or environment, reflections on whether interview questions are working well, connections between the responses that multiple interviewees gave you, and so on. We suggest that as soon after your interview as possible, you conduct a brain dump. Use any jottings you wrote during or immediately after each interview to draft a formal field note that contains key quotes, descriptions of what topics you discussed, and any observational notes, including any details you think are relevant about the setting where the interview was conducted or about your interviewee's appearance and mannerisms.

Sometimes qualitative sociologists create documents separate from field notes—called *journals* or *memos*—where they put personal reflections and analytical insights that come to them during the data collection process. You might just incorporate these preliminary thoughts about your research into your field notes, but regardless, you should write them down somewhere while you are still gathering data. For example, an interviewee might suggest a new concept you hadn't thought of before or define a concept in a new way. This may lead you to create new questions or ask questions in a different way to future interviewees. When writing down notes between interviews, you can reflect on your own feelings about your research and talk through any methodological decisions you made. Documenting these details is important, as you will be surprised by how quickly you will forget what happened during your fieldwork. When the time comes to analyze your data, your notes will help you remember how, when, and why you made certain decisions or made certain connections. These notes also provide evidence of the rigor of your research process, including its trustworthiness and authenticity (refer to our discussion of these standards of qualitative rigor in **Chapter 7: Measuring the Social World**).

Key Takeaways

1. To recruit interviewees, qualitative researchers tap network-based referrals and pursue offline and online means of publicizing their studies to potential volunteers.
2. Conducting interviews in people's homes or workplaces can provide useful context, though those advantages need to be weighed against the comfort and security of both the interviewer and interviewee.
3. Qualitative researchers should take detailed field notes to summarize their interviews, record nonverbal details not available in transcripts, and capture their in-process reflections about their research.

10.4. Conducting In-Depth Interviews

Learning Objectives

1. Describe the importance of rapport in an in-depth interview and how it can be cultivated.
2. Discuss strategies for getting interviewees to elaborate on what they say.
3. Describe the power differentials inherent in qualitative interviewing and how researchers might address them.

A unique feature of interviews is that they require some social interaction, which means that a relationship is formed between the interviewer and interviewee. **Rapport** is the sense of connection you establish with a research participant. One strength of in-depth interviewing, as we've mentioned, is that it generates a deep understanding of how people think about and experience their social realities. It can do so because of the bond of trust that the interviewer and interviewee develop over the course of their interactions. Building rapport makes the interviewee more likely to answer your questions with care and candor.

How do you cultivate rapport with your interviewees? A central component of rapport is respect. In-depth interviews are intimate encounters. Your participants will share with you how they view the world, how they understand themselves, and how they have coped with events that have happened to them. You should show respect for their story by not judging them. You are there to listen to them. They have been kind enough to give you their time and attention. Even if you disagree strongly with what an interviewee tells you in an interview, your job as the researcher is to learn what they know and feel, not to criticize them. This is not to suggest that you can't push back on what your interviewees say—doing so is sometimes necessary to highlight contradictions or tensions in their thinking—but you should do so respectfully, without disdain or a sense of superiority, and you should be judicious about when and how often you choose to challenge people. It is also crucial that you, as the interviewer, conduct the interview in a way that is culturally sensitive. In some cases, this might mean educating yourself about your study population and even receiving some training to help you learn to effectively communicate with your research participants.

Developing good rapport requires good listening. In fact, listening during an interview is an active, not a passive, practice. You should engage with your interviewees by showing you understand whatever it is that they are telling you (Devault 1990). The follow-up questions you ask them should indicate you've actually heard what they've just said. You can also indicate that you are following along by giving verbal and nonverbal cues—for instance, nodding your head when you agree with what the interviewee said, or saying words like, "hmm," "wow," or "interesting." Obviously, you can overdo this sort of encouragement, and you should not fake your enthusiasm or agreement—which, besides being duplicitous, is something many interviewees can readily sense.

In general, you should shoot for a natural flow of questions and answers in your interview, much like the back-and-forth of a normal conversation. Avoid just following the script of the interview guide, which can make the interview seem overly formal and forced. Instead, deepen the conversation by using probes—either ones included in your interview guide or improvised on the spot. This will strengthen your rapport with the interviewee. With time, they will come to see you as someone who truly cares about what they have to say.

As noted earlier, however, it is important to recognize that an interview is not a normal conversation: it is a controlled

conversation. It is intended to be one-sided—which means that you as the researcher should not be doing much talking. In some situations, it may be useful to share something about your own personal views or background, but for the most part, you need to let the interviewee talk, only interjecting when you want to encourage them to say more. We can't stress enough how important it is to know when to shut up during an interview. When you listen to your interview recordings afterward, you may be surprised by how often you talk over an interviewee just when they are about to disclose something personal or say something profound—try to curb this natural desire to weigh in.

As the interviewer, you want to control where the conversation goes. You probably don't want the interviewee to go off on a 15-minute rant on their gripes about the current phase of the Marvel Cinematic Universe if that isn't at least tangentially related to your research question. Yet abruptly cutting them off can make you come across as rude, hurting your ability to develop rapport with the interviewee. Consider using gentler nudges—for instance, "Let's get into that a bit later," or "What you said reminds me of . . ."—to bring them back to subjects you care about. That said, if an interviewee is particularly excited about discussing something, and it has something to do with your research question, you might want to hold off on pivoting to another topic. As we've noted, one of the key strengths of qualitative research is its ability to surprise you as a researcher, so you don't necessarily want to cut off an interviewee who's on a roll: it may lead your conversation in an unexpected but fruitful direction.

On a related note, we would strongly advise you not to rush an interviewee if that person is sharing something deeply personal with you. In fact, you should show gratitude for this gesture of trust in you by being silent and listening intently. You want to make sure the interviewee feels heard, and showing respect for a person's story is critical. If the interviewee is expressing strong emotions, you can offer comfort and encouragement whenever you feel it is both welcome and appropriate. The interviewee should not be treated as a means to an end, and in this situation, the interviewer has a moral responsibility to show empathy and compassion and absorb what the other person has to say without judgment. Always remember that you are a human being first, and a researcher second.

Getting the Most Out of Your Questions

When you are actually conducting an in-depth interview, perhaps the most important thing to keep in mind is asking follow-up questions. As we noted, this shows that you are actively listening to your interviewee, strengthening your rapport with them. It also helps you generate the thick description that is essential to qualitative work.

A key strategy in probing for more information is to get the interviewee to "show, not tell" (an adage that exemplifies the approach of qualitative research more broadly, as we noted in **Chapter 9: Ethnography**). In in-depth interviewing, this means getting your interviewee to give you specific examples to back up whatever statements or judgments they are making. Here are some examples of probes you can use to get your interviewee to provide examples:

- What would be an example of that?
- When you say that, what are you thinking of?
- Can you tell me the last time that happened?
- Can you think of a time that really showed that happening?
- Can you walk me through what happened?

You may also want to urge the interviewee to be detailed in describing whatever examples they give you—perhaps by saying something like, "The more concrete you can be, the better" (Weiss 2004).

The worst-case scenario for a qualitative researcher is an interviewee who has very little to say in response to your carefully crafted questions. Sometimes, there is nothing you can do about that: the interviewee just doesn't want to talk, and you'll have to stumble painfully through your interview with them, asking question after question to a blank wall of silence.

Nevertheless, you can employ several simple strategies to get people to talk in greater depth. One is to use stock probes—catchall follow-up questions that you can slip into the conversation to get an interviewee to elaborate on what they just said. Here are some examples:

- Why? (*Or*: Why do you say that?)
- How is that? (*Or*: How do you mean that?)
- In what ways?
- Could you tell me a little more about that? (*Or*: Could you elaborate a bit on that?)

Depending on how or when you use them, stock probes like "Why?" can come off as confrontational, even if that is not your intent. Oftentimes people won't know how to respond to "Why?" in particular, perhaps because they don't even know why they feel that way. If you're worried about coming across as doubting or challenging, you might want to use the last option in our list (or any variation of "say more about that," really), which provides a gentler way of encouraging interviewees to explain themselves. Alternatively, you could use more specific probes, asking for clarification or confirmation about what the person just said. These might be yes/no questions like, "So that friend you mentioned didn't agree with your decision?" Or, you might pose simple questions asking for further details like, "When did that happen?" The interviewee may respond in just a word or two, but this line of questioning may eventually prompt a person who stopped talking to pick up the topic again.

To deepen the conversation, you can also follow the tried-and-true active listening strategy of repeating what was just said in your own words. You could summarize what you've heard, followed up by a question like, "Is that you mean?" Alternatively, you could slip in a brief statement that affirms or comments on what was just said. For instance, if an interviewee has just narrated an intense argument that their parents had when they were a child, you might say something like, "It sounds like your parents had a difficult relationship"; if they are talking about all the loans they have taken out to pay for college, you might want to comment on how personally challenging it was for them to pursue this goal. This active-listening approach not only helps you stay on top of what the interviewee is saying, but your interjection might encourage them to say more on the subject at hand—even if just to disagree with your summary or comment.

Asking follow-up questions is essential in in-depth interviewing, but the best way to get the interviewee to say more is often to just stay quiet. As National Public Radio host Rachel Martin said about conducting journalistic interviews: "Sometimes the best follow-up question is to say nothing. Especially after the person has just revealed something important. More often than not, if you are quiet and give them the space to keep sharing—they will." The same principle applies in sociological research. While you don't necessarily want too many awkward silences in an interview (remember the importance of building rapport!), don't try to fill in every pause in the conversation with your own words. Sometimes silence is the most effective probe.

Beyond continually probing for deeper information, the following strategies can make your interviews more productive:



Resist the urge to fill the silences of an interview. By just pausing for a moment before you pose a follow-up question or turn to another topic, you give your interviewee a chance to collect their thoughts and articulate a more complex or controversial view. Letting the silence sit for a moment will naturally encourage the interviewee to step in and add to what they just said. (Lots of silence can be awkward, though, so don't overdo it!) *Jono Hey, via* Sketchplanations

1. **Don't be afraid to ask dumb questions.** When people are talking about topics they know a lot about, they may assume you are an insider who knows all the lingo and context. Obviously, you should do your homework before the interview to understand as much about any relevant topics as you can, but if you don't understand something, say so: you don't want to nod along amiably while steadily growing more confused. You'll find that most of the time, people are happy to explain things to someone new to their area of expertise.
2. **Move regularly between "miner" and "traveler" questions during your interview.** A "miner" delves deeply into a specific area; a "traveler" ranges about the terrain, covering a wide range of topics (Kvale 1996). You want to find a nice balance between these two modes of asking questions. Sometimes you will want to gather meticulous details about a phenomenon, nudging the interviewee to give you concrete examples of it and describe it in all its complexity. Other times, you will want to move quickly between distinct topics—thereby ensuring you have covered everything relevant—or take the conversation to a higher level—getting your interviewee to think of the big picture and share their thoughts, in broad brushstrokes, on an issue or set of issues. Both approaches are essential to conducting a good qualitative interview.
3. **Stay focused for the entirety of the interview.** Building rapport takes time, and the later parts of the interview (or later interviews with the same person) will frequently feature more candid and explicit discussion. Until the interviewee is comfortable with you, they might give you formulaic or superficial answers to your questions—especially for any that have a hint of stigma or controversy attached to them. In fact, you may want to return to these difficult questions later in an interview, after the interviewee has warmed up to you, to see if they have become more open to answering them or providing additional details.

Power Dynamics in an In-Depth Interview

Interviewers must be attentive to the inevitable power differential between themselves and the people they interview. This is one example of the **reflexivity** that qualitative researchers engage in whenever they go into the field. As ethical researchers, we should seek to understand our social position in relation to those we are studying (our **positionality**) as well as the impact we can have on them—positive and negative—in part because of the underlying inequalities in our relationships.

As we noted, the interview is a controlled conversation, and it is the interviewer who sets the agenda and leads the conversation. Qualitative interviewers aim to allow participants to have some control over which topics are discussed, but at the end of the day, it is the researcher who is in charge of the interview and how any of its findings are reported to

the public. The interviewee loses the ability to shape the narrative after the interview is over because it is the researcher who tells the story to the world.

In a good in-depth interview, interviewees may reveal things about themselves they may not typically share with others. Yet researchers do not reciprocate by revealing much or anything about themselves. As much as we have emphasized the importance of rapport in in-depth interviewing, it is important to consider whose interests are being served by this rapport. In fact, some critics see rapport as a ruse—the interviewer tricks the interviewee into believing they are closer to the interviewer than they really are, which means they divulge sensitive information at their expense but the interviewer's benefit (for a discussion of these issues, see the sidebar **The Journalist and the Ethnographer** in **Chapter 8: Ethics**).



Qualitative researchers working from a feminist tradition have argued that in-depth interviewers should allow their respondents to review interview transcripts before analysis and paper drafts before publication. The idea is that this collaborative approach makes the relationship between the interviewer and interviewee more reciprocal and reduces the power imbalance inherent to this research method. *Tim Lauer, via Flickr*

As we talked about in our earlier discussion of ethics, one of the ways that researchers address these power differentials is by providing as much information to research participants as possible so they can make informed decisions about whether to be part of a study. As part of that consent process, we ensure that interviewees understand very clearly the purpose and approach of our research. We share with them how the data we gather will be used and stored. And we explain what procedures will be taken to protect their identities and minimize any other risks of the study. Some researchers have argued, however, that qualitative researchers should go further in addressing power imbalances between interviewers and interviewees. Feminist researchers in particular have pioneered strategies such as making interviews a more reciprocal exchange of experiences, allowing participants to view and edit interview transcripts before the researcher uses them for analysis, and giving participants an opportunity to read and comment on papers before they are published (Hesse-Biber 2013; Oakley 1981; Reinharz 1992). That said, sociologists disagree about what level of involvement research participants should have in a qualitative study's data collection and analysis, and even these suggestions have been criticized as conveying a false sense of security, when in reality researchers can still present participants' stories in whatever way they see fit (Stacey 1988).

Ultimately, there are no easy answers when it comes to addressing the power differential between interviewers and interviewees. Because qualitative research involves engaging in interpersonal interactions and building relationships with research participants, there is always a danger that researchers will use the rapport they cultivate to take advantage of those they study. You as a sociologist will need to continually reflect on your relationships with your interviewees and figure out the best way to safeguard their well-being while successfully carrying out your research.

Capturing Stories that Resonate: A Q&A with Allison Pugh



Allison Pugh is a professor of sociology at the University of Virginia and the chair of its Department of Women, Gender, and Sexuality. Her research and teaching focus on how economic trends—from job insecurity, to commodification, to automation—shape the way people forge connections and find meaning and dignity at home and at work. She has also written about the theoretical benefits and best practices of interpretive interviewing, and about resonance in ethnography—when stories of other people's lives make powerful meaning for a larger world (Pugh 2013, 2014).

Drawing from extensive in-depth interviews, Pugh's acclaimed book *The Tumbleweed Society: Working and Caring in an Age of Insecurity* (2015) is a study of the broader impacts of job precariousness. Another of her books, *Longing and Belonging: Parents, Children, and Consumer Culture* (2009), received four honors from the American Sociological Association and was a finalist for the C. Wright Mills Award. Her current research, funded by the National Science Foundation, is a study of the standardization of work that relies on relationships. She has been a fellow of the American Council of Learned Societies, the Berggruen Institute, and the Center for Advanced Study in the Behavioral Sciences at Stanford, and a visiting scholar in Germany, France, and Australia.

How did you get into sociology?

I found sociology relatively late, after I had already been a journalist, a diplomat, and a peon in advertising. In

1994, I enrolled as a "community scholar" at Berkeley. That's the way that Berkeley lets people from the community—that is, not regularly enrolled students—take classes. I took three classes in the fall and three classes in the spring before I actually enrolled as a graduate student. I ended up taking a family sociology class with Arlie Hochschild, a feminist theory class with Nancy Chodorow, and a cultural sociology class with Ann Swidler—all three were incredible. I thought I was just looking for a place that would give me structure while I wrote a nonfiction book, but I read Arlie's book *The Second Shift* and was hooked.[1]

How do you use in-depth interviews in your own research?

I rely on them as the primary prism through which we can view how people experience and interpret their own sorrows and triumphs. They are a source of inspiration and of exhaustion. I often am particularly attuned to sadness or compromise, the evidence of all the ways we struggle to reconcile ourselves to the contradictory dictates of social life—for example, to be an ideal worker and an ideal parent. I also use them as a guide, a flag of sorts for what we might be missing or misunderstanding in scholarship. Each one is a gift from the interviewee.

You've interviewed a wide range of people over the course of your projects—from parents and children in *Longing and Belonging*, to diverse kinds of workers (including the unemployed) in *The Tumbleweed Society*. Do you find yourself taking a different approach to your interviewing when you talk with different sorts of people? Are there any particular concerns or strategies you consider when interviewing people from one group over another?

I suppose my first concern is always rapport, and how to achieve that can vary depending on the interviewee. There are basics, of course. Try to make it as natural a conversation as possible—avoid reading lengthy questions from long lists on paper, for example. Pay attention to the length of their answers as well as to their body language. If there are rapport problems, I sometimes abandon my whole interview schedule and just try to learn more about them—essentially, following them where they want to go—before figuring out how to ask my questions of interest. I might also change my own presentation of self, depending on who I am meeting. I might dress up a little more for some formal settings, but be more casual in other settings.

But at its core, rapport is about how much you listen to and see the other person, and reflect that back to them, so they know they are seen—that is the path to a good interview, and more important than any other practice. That is a constant across interviewees.

What advantages do you think in-depth interviews offer as a research method over other approaches?

At the most fundamental level, it allows us to hear how people narrate their own lives, struggles, issues. But as they tell their stories, the very language they choose—the metaphors, the emotional descriptors, etc.—also offers us very compelling data on what they actually think, not just what they think they are supposed to think. Interviews allow us to understand what it feels like to live in their lives for a while. Other methods are better at investigating the extent of an issue or phenomenon, or, as in ethnography, documenting the practical intersections between an individual's views and the social world in which they must translate them into action. But interviewing gives us someone's own perspective—their own commitments and compromises—like no other.

You are very skilled at getting your interviewees to talk about intimate details of their lives. How has your experience been in getting people to open up about sensitive subjects and confide in you? What advice would you give novice researchers about how to develop rapport with their respondents?

1. Hochschild (1989).

Successful rapport, and interviewing in general, relies almost entirely on listening really intently, hearing the other person and reflecting that back to them. What you are listening for is not just what they are saying but, underneath that, what—or what else—they really mean. If, for example, someone tells you they are proud of their son for standing up to his school, but they also communicate worry—they fidget a lot, or they can't seem to let it go somehow, repeating themselves or talking a lot about it, or they say they are proud but their eyebrows are high and their forehead is furrowed—I might say something like, "It sounds like you are proud of him, but I am also hearing a lot of worry there, too."

That can feel like taking a risk, especially for new interviewees—*what if I'm wrong*? In fact, it is okay to be wrong—they'll correct you, it's a natural conversation. But they usually appreciate your effort to get it right, and if you get it right, and someone feels seen on both a surface and a deeper level, it can actually offer some relief to them. Naming the unnamed, carefully, in a way that doesn't threaten them unduly, is the best way to help someone feel safe in disclosing their inner thoughts or feelings.

Have you come across any challenges balancing the researcher's need to present an accurate perspective of an interviewee's feelings, beliefs, and actions and our ethical obligations to be humane and respectful about how people are portrayed in print? How do you balance these goals in your own work?

I suppose I approach my interviewees as if they are not that different from me. My goal is to understand how they think and feel, and to reckon with how their perspectives and practices make sense in their environment. We are all just trying to get dignity and respect in our own social worlds. If we remember that, it makes their particular foibles or flaws understandable. I am not sure I would do "better than" any person I have interviewed if I were given their origins and experiences, and that's true for some pretty tough folks. In my research for *Tumbleweed*, I heard people repeatedly invoking respected others—their priest, their brother, their therapist—when they explained why they had to abandon or detach themselves from family or friends. I did not think to myself, "Well, I'm sure I would not do that." Instead, I thought, "What is the underlying conflict, the cultural contradiction, even the desperation, that is leading people to turn their backs on people they love?" It's not useful or interesting or even true to point to "good" or "bad" people. I consider it my job to name the cultural contradictions that are pressing upon people, the systems that generate the conditions of people's choices.

Where do you personally fall in some of the recent debates about transparency in interview-based research—for instance, whether sociologists should use the real names of their respondents when given permission, whether they should offer their data for public scrutiny—as many journals now require for quantitative data—and whether they should independently fact-check the statements of respondents?

I have a piece, written with a former student, on this very issue.[2] I feel very strongly that we owe respondents confidentiality; that they do not always understand what it will feel like to be the subject of a book or article at the time of the research; and that concealing their identities gives them what Victoria Reyes has called "plausible deniability."[3] Protecting them in this way feels like the least we can do, in the face of the enormous gift of their participation in our research. Not protecting them exposes them to potentially vituperative social opprobrium—in other words, trolls—and saying they consented to it just deflects the blame from us, as an ethical fig leaf, as it were. Meanwhile, fact-checking fetishizes "facts" as if they are the source of trust in

2. Pugh and Mosseri (2023).

3. Reyes (2018).

qualitative methods. If someone does not trust the book or article after they have read it, "fact-checking" is not going to change their mind. Furthermore, as Sarah Mosseri and I write in our article, many of the proposed reforms of qualitative research actually threaten the very aspect that makes it precious: its capacity for resonance.

There have been incidents in the past where sociologists have been called out about the veracity or ethics of their qualitative methods. What do you think qualitative sociologists can do to inspire confidence in their approaches?

Qualitative researchers have longstanding practices of reflexivity that incrementally build trust in their work. This includes being frank about our positionality, being detailed about our processes of data collection and analysis, and writing explicitly about exceptions, "negative cases," and variability. But ultimately, qualitative research—like many good things in the world—does rely on a modicum of trust, which I do not regard as a flaw. Qualitative sociology also gets read by a large audience, and has been for more than a century. I would argue there is already substantial confidence in its approaches, despite a few recent controversies.

What advice would you give an aspiring sociologist if they wanted to pursue interview-based research?

Interviewing is widely used as a method in sociology, so I would not be worried about them or their future careers. I would advise them to practice! Interviewing relies on oral and interpersonal skills, and many aspiring sociologists are good at reading, writing, critiquing arguments, etc., but less so at reading another person's signals, managing the pace of an interview, and handling emotional clues. These are skills you can develop. Also, I would advise them to read up on the debates and advances in the field so that they can be intentional about the choices they are making. There has been a lot of good work in qualitative methods recently, from Mario Small and Kristin Luker's advice about research design and how to devise a sample, to Nicole Deterding and Mary Waters's suggestions for effective coding, to Iddo Tavory and Stefan Timmermans's work on abductive analysis.[4] While I still regard Robert Weiss's *Learning from Strangers* as a vital guide, there are two strong recent books by Kathleen Gerson and Sarah Damaske and Annette Lareau, and some more in the offing, as I understand.[5] It's a good time to be a qualitative researcher.

Key Takeaways

1. By building rapport with interviewees, a researcher is able to maximize the ability of in-depth interviews to obtain detailed and intimate knowledge of an individual's experiences and attitudes.
2. Interviewers can use probes—both planned and improvised—to gather concrete examples and encourage interviewees to go into greater detail.
3. Interviewers should recognize the power differential between themselves and their interviewees and try

4. Small (2009); Luker (2010); Deterding and Waters (2021); Tavory and Timmermans (2014).

5. Weiss (1995); Gerson and Damaske (2020); Lareau (2021).

their best to minimize it.

10.5. Focus Groups

Learning Objectives

1. Describe how focus groups operate and how they differ from one-on-one interviews.
2. Describe the considerations that go into determining the size and composition of focus groups.
3. Discuss the various roles that researchers play when running focus groups.
4. Identify the strengths and weaknesses of focus groups as a research method.

Focus groups are qualitative interviews focused on a particular topic that researchers conduct with multiple respondents at the same time. They resemble one-on-one in-depth interviews in that a researcher may prepare an interview guide in advance and interact with participants by asking them questions in a more or less unstructured fashion. But anyone who has conducted both one-on-one interviews and focus groups knows that each is unique. In a one-on-one in-depth interview, usually one person (the interviewee) is most active while the other (the researcher) plays the role of listener, conversation-guider, and question-asker. Focus groups, on the other hand, are meant to spark discussion and debate within the group. As a result, researchers play a very different role than they do in one-on-one interviews. Rather than just asking questions, they are actively trying to get participants to talk to each other. They observe these interactions among participants in addition to listening to what each person has to say.

Focus groups are popular among researchers outside academia. One of the earliest documented uses of focus groups comes from World War II, when government researchers used them to assess the effectiveness of troop training materials and various propaganda efforts (Morgan 1997). Market researchers quickly adopted this method to gather information about the products or services they aim to sell. Today, focus groups are widely used by marketers as well as researchers working on behalf of government agencies and political campaigns, who often employ them to learn how members of the public feel about a particular issue or candidate. Within the social sciences, the use of focus groups did not really take off until the 1980s, when demographers and communication researchers began to appreciate their use in understanding knowledge, attitudes, and communication.

In the following sections, we'll describe some of the practical considerations that go into planning and implementing focus groups.

Who Should Be in Your Focus Group?

A focus group can include anywhere between 3 and 12 participants (Adler and Clark 2007). One key consideration when determining the size of a focus group is the topic of the interview—specifically, whether it is a subject that participants will have a lot to say about without much prompting. If participants are likely to feel passionately about the topic being discussed, a group of three to five makes sense. Groups larger than that, especially for heated topics, can easily become unmanageable. If you are new to running a focus group, you may also want to keep the group on the smaller side.

Focus group researchers must carefully consider the composition of the groups they put together. The background characteristics of participants—their age, gender, race, class, sexuality, and so on—will influence the extent to which specific individuals are silent or talkative during the conversation, as well as what they are willing to say in front of others

(Hollander 2004). It may seem counterintuitive, but it is usually better to form focus groups consisting of participants who do not know one another (Agar and MacDonald 1995). Friends, relatives, or acquaintances may share some taken-for-granted knowledge, and it is often these sorts of assumptions that researchers are looking to uncover. Focus group researchers should therefore avoid setting up interactions where participants may be discouraged to question or raise issues that they take for granted. However, members of the group should not be so different from one another that participants do not feel at ease talking to one another. Focus group researcher David Morgan (1997:36) suggests that "homogeneity in background and not homogeneity in attitudes" should be the goal, since participants must feel comfortable speaking up but must also have enough differences to facilitate a productive discussion.

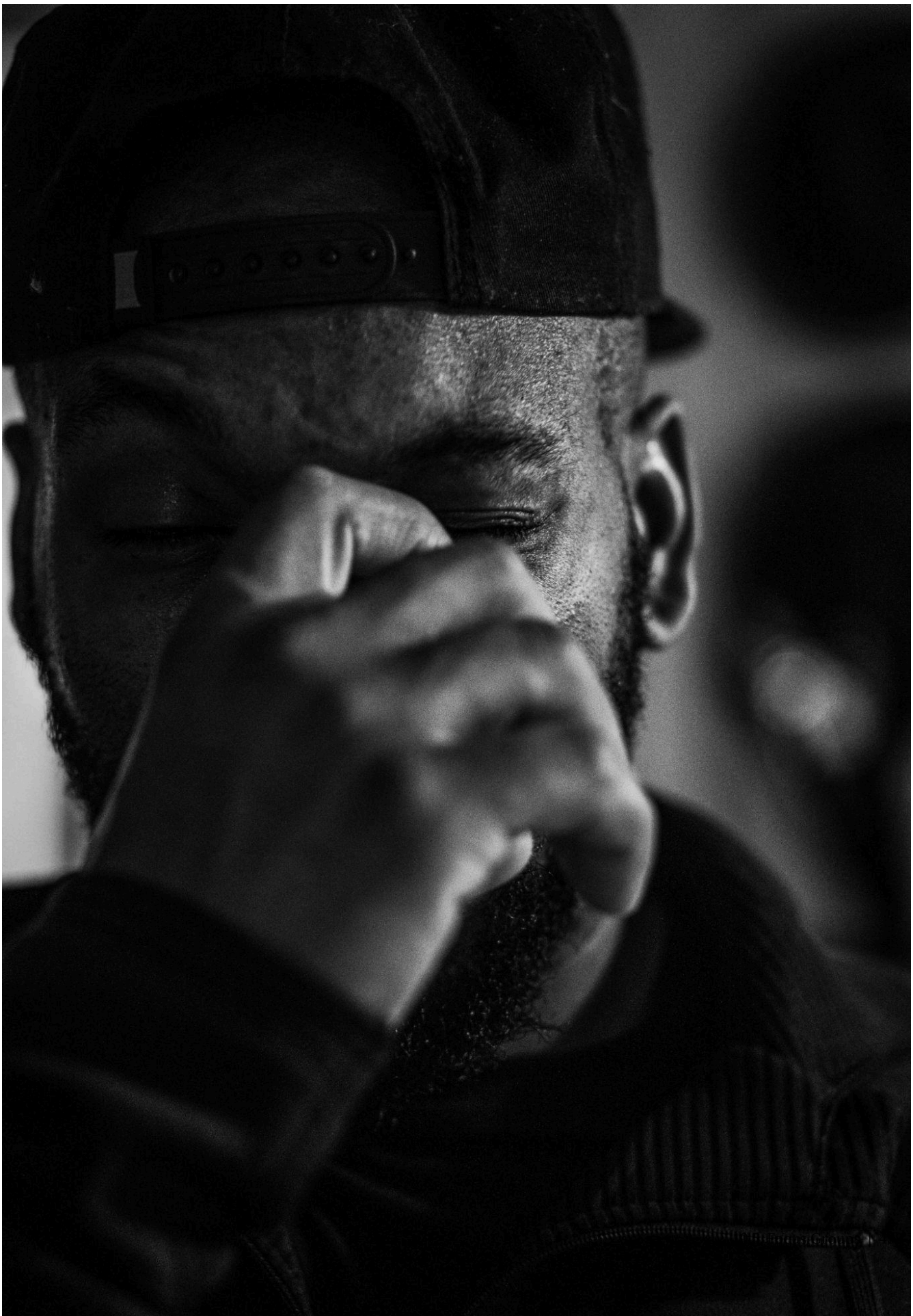
The Role of the Moderator

In a focus group, participants may ask each other follow-up questions, agree or disagree with one another, and even come up with questions not previously conceived of by the researcher. All of these details are potentially useful—in addition, of course, to each individual's direct responses to the questions posed. Ensuring fruitful interactions within the focus group is the responsibility of the **moderator** (also known as a *facilitator*), a researcher tasked with overseeing the conversation in a focus group. The moderator's job is not to ask questions to each person individually, but to stimulate conversation between participants.

When you moderate a focus group, you should set ground rules at the outset of the discussion. Ask participants to avoid having side conversations; thoughts or reactions to what is said in the group are important and should be shared with everyone. Remind participants you've invited them to the group because you want to hear from all of them. With that goal in mind, only one person should speak at a time, and a few participants should not dominate the conversation. To set a tone of participation by everyone, you can begin the discussion by asking each participant to introduce themselves or provide a response to an opening question. Although there are risks that certain members of a focus group will monopolize the conversation, it's worth noting that focus groups are also good at getting people to participate who may be intimidated by a one-on-one interview setting because they feel they have nothing to contribute. In a focus group, the encouragement and moral support provided by other group members may encourage them to speak up (Kitzinger 1995).

The moderator will play a less active role as the focus group starts chugging along and participants start talking to one another. That said, there may be times when the conversation drags or when you, as moderator, wish to guide the conversation in another direction. As we discussed with one-on-one interviews, it is best not to abruptly shift gears. Show that you have been listening carefully to what people have been saying with a deft transition: "So several of you have mentioned X. But what about Y?" Likewise, to shift the focus to participants who have been less active in the group, you might say something like: "Let's hear more about what Delaney and Sunil think about what Dominick has been saying."

Researchers may choose to use multiple moderators to make managing these various tasks easier. They may record the focus group session. And if moderators have their hands full with running the meeting, they may recruit a notetaker to draft field notes based on what transpired during the focus group. In addition to any patterns or themes observed in people's responses, the notetaker can record what expressions, mannerisms, and other nonverbal cues people used to convey their feelings, how the participants influenced the flow of the conversation through moments of humor or interruptions, and how they managed disagreements (Liamputtong 2011). Focus groups are analyzed in a similar way as one-on-one interviews are, but this interactive dimension between participants—hopefully captured by the notetaker and any video or audio recordings—adds additional depth to the data collected.



You can observe how people react when they are answering interview questions or engaging in a focus-group conversation. Their facial expressions and body language can give you valuable clues about how they feel about a particular topic—even how they might have mixed or contradictory feelings about what they are telling you. Catching these nonverbal cues can be all the more helpful when conducting focus groups, which feature interactions between group members that provide ample amounts of this observational data. *Marcus Bellamy*, *via* *Unsplash*

Strengths and Weaknesses of Focus Groups

Focus groups share many of the strengths and weaknesses of one-on-one qualitative interviews. Both methods can yield very detailed, in-depth information. Both are excellent for studying social processes. And both provide researchers with an opportunity not only to hear what participants say but also to observe other cues about their feelings, attitudes, and backgrounds. Focus groups offer the added benefit of giving researchers a chance to collect data on human interaction by observing how group participants respond and react to one another.

Like one-on-one qualitative interviews, focus groups suffer from the drawback of requiring a substantial amount of time and effort on the part of researchers. In fact, focus groups arguably require more planning than any other qualitative method of data collection. In one-on-one interviews, for example, researchers can improvise questions based on how the conversation unfolds. But focus groups must be carefully constructed from the beginning. As we noted, researchers need to recruit participants who will comfortably interact with one another. They also need to control the timing of the event so that participants are not asked to stay for a longer time than they've agreed to participate. That said, there may be some time savings with focus groups as it takes fewer group events than one-on-one interviews to gather data from the same number of people.

A weakness that is specific to focus groups is that one or two participants might dominate the conversation, silencing other participants. Similarly, a focus group whose members do not work well with one another can wind up in a state of disengagement or hostility, making it hard to generate the stimulating, interactive discussions that are the hallmark of an effective focus group. Addressing these potential problems requires researchers to be careful in selecting the members of their focus group and skilled in moderating the session in real time. Another drawback of focus groups is that they may worsen problems with social desirability bias in qualitative interviews, given that participants are being judged by a larger audience. That said, a sense of camaraderie within a focus group may also encourage reticent members to speak up, and having one or more members broach sensitive topics can prompt others to share their views or experiences. A more intractable problem for focus groups is that they weaken the ability of researchers to protect the privacy of a study's participants. While it is standard for researchers to tell participants to keep what is said in the group confidential, they cannot control what individuals say to other people after the focus group has wrapped up.

Key Takeaways

1. In contrast to one-on-one interviews, the goal of a focus group is to get participants to talk with one another, a conversation that the researcher moderates.
2. A focus group can have anywhere from 3 to 12 members, though a smaller number is better for topics that people have a lot to say about. Ideally, the members of a focus group will have a relatively homogeneous background (ensuring an adequate degree of comfort in their interactions) but also represent a diversity of opinions (ensuring a vigorous discussion).

3. Focus groups can provide additional details that one-on-one interviews cannot regarding the social interactions between participants as they discuss and debate the issue at hand, possibly building on each other's ideas.
4. Focus groups require special planning and suffer from distinct problems, such as difficulties in ensuring confidentiality and adequate participation by everyone in the group.

10.6. References

Adler, Emily Stier, and Roger Clark. 2007. *How It's Done: An Invitation to Social Research*. 3rd ed. Belmont, CA: Wadsworth.

Agar, Michael, and James MacDonald. 1995. "Focus Groups and Ethnography." *Human Organization* 54(1):78–86.

Chen, Victor Tan. 2015. *Cut Loose: Jobless and Hopeless in an Unfair Economy*. Oakland: University of California Press.

Deterding, Nicole M., and Mary C. Waters. 2021. "Flexible Coding of In-Depth Interviews: A Twenty-First-Century Approach." *Sociological Methods and Research* 50(2):708–39. doi: 10.1177/0049124118799377.

Devault, Marjorie L. 1990. "Talking and Listening from Women's Standpoint: Feminist Strategies for Interviewing and Analysis." *Social Problems* 37(1):96–116. doi: 10.2307/800797.

Esterberg, Kristin G. 2001. *Qualitative Methods in Social Research*. Boston: McGraw-Hill.

Gerson, Kathleen, and Sarah Damaske. 2020. *The Science and Art of Interviewing*. New York: Oxford University Press.

Hesse-Biber, Sharlene Nagy, ed. 2013. *Feminist Research Practice: A Primer*. 2nd ed. Thousand Oaks, CA: SAGE.

Hochschild, Arlie. 1989. *The Second Shift*. New York: Viking Adult.

Hollander, Jocelyn A. 2004. "The Social Contexts of Focus Groups." *Journal of Contemporary Ethnography* 33(5):602–37. doi: 10.1177/0891241604266988.

Kitzinger, Jenny. 1995. "Qualitative Research. Introducing Focus Groups." *BMJ* 311(7000):299–302. doi: 10.1136/bmj.311.7000.299.

Kvale, Steinar. 1996. *InterViews: An Introduction to Qualitative Research Interviewing*. Thousand Oaks, CA: Sage.

Lareau, Annette. 2021. *Listening to People: A Practical Guide to Interviewing, Participant Observation, Data Analysis, and Writing It All Up*. Chicago: University of Chicago Press.

Liamputtong, Pranee. 2011. *Focus Group Methodology: Principles and Practice*. Thousand Oaks, CA: SAGE.

Luker, Kristin. 2010. *Salsa Dancing into the Social Sciences: Research in an Age of Info-Glut*. Cambridge, MA: Harvard University Press.

McAdam, Doug. 1990. *Freedom Summer*. New York: Oxford University Press.

Morgan, David L. 1997. *Focus Groups as Qualitative Research*. 2nd ed. Thousand Oaks, CA: SAGE.

Oakley, Ann. 1981. "Interviewing Women: A Contradiction in Terms." Pp. 30–61 in *Doing Feminist Research*, edited by H. Roberts. London: Routledge and Kegan Paul.

Pugh, Allison. 2009. *Longing and Belonging: Parents, Children, and Consumer Culture*. Berkeley: University of California Press.

Pugh, Allison J. 2013. "What Good Are Interviews for Thinking about Culture? Demystifying Interpretive Analysis." *American Journal of Cultural Sociology* 1(1):42–68. doi: 10.1057/ajcs.2012.4.

Pugh, Allison J. 2014. "The Divining Rod of Talk: Emotions, Contradictions and the Limits of Research." *American Journal of Cultural Sociology* 2(1):159–63.

Pugh, Allison J. 2015. *The Tumbleweed Society: Working and Caring in an Age of Insecurity*. New York: Oxford University Press.

Pugh, Allison, and Sarah Mosseri. 2023. "Trust-Building vs. 'Just Trust Me': Reflexivity and Resonance in Ethnography." *Frontiers in Sociology* 8:1069305. doi: 10.3389/fsoc.2023.1069305.

Reinharz, Shulamit. 1992. *Feminist Methods in Social Research*. New York: Oxford University Press.

Reyes, Victoria. 2018. "Three Models of Transparency in Ethnographic Research: Naming Places, Naming People, and Sharing Data." *Ethnography* 19(2):204–26. doi: 10.1177/1466138117733754.

Small, Mario Luis. 2009. "'How Many Cases Do I Need?': On Science and the Logic of Case Selection in Field-Based Research." *Ethnography* 10(1):5–38. doi: 10.1177/1466138108099586.

Stacey, Judith. 1988. "Can There Be a Feminist Ethnography?" *Women's Studies International Forum* 11(1):21–27. doi: 10.1016/0277-5395(88)90004-0.

Tavory, Iddo, and Stefan Timmermans. 2014. *Abductive Analysis: Theorizing Qualitative Research*. Chicago: University of Chicago Press.

Weiss, Robert S. 1995. *Learning From Strangers: The Art and Method of Qualitative Interview Studies*. New York: Free Press.

Weiss, Robert S. 2004. "In Their Own Words: Making the Most of Qualitative Interviews." *Contexts* 3(4):44–51. doi: 10.1525/ctx.2004.3.4.44.

11. QUALITATIVE DATA ANALYSIS



How can you analyze a vast set of empirical evidence acquired through participant observation, in-depth interviews, focus groups, audio or video recordings, or secondary documents? Qualitative data analysis specializes in condensing large amounts of qualitative data into more manageable bits of information that represent patterns and relationships visible across a sample. During this analysis process, you review your ethnographic field notes, interview transcripts, or other qualitative data sources and try to make sense of what is there: what tendencies can be seen in the data, what processes are described or explained, what social meanings are being conveyed by people's actions and words, and so on.

How is qualitative data analysis different from quantitative data analysis? To review what we discussed in **Chapter 4: Research Questions**, experiments and survey research typically take a deductive approach that is focused on theory testing—providing evidence to support or challenge researchers' hypotheses. Qualitative research, in contrast, tends to pursue an **inductive approach** that is about theory building. While the deductive researcher starts with a theory, derives concrete hypotheses from that theory, and tests those hypotheses using empirical data, the inductive researcher starts with data and tries to sketch out a theory that makes sense of the patterns in that data.

The fact that a qualitative study often starts with immersing yourself in the data is important. That's because one key advantage of the qualitative approach to data collection and analysis is that it allows for these two procedures to occur more or less at the same time. For instance, you might decide to conduct an interview and do a preliminary analysis of your data before proceeding to the next interview. Doing so allows you to correct potential flaws in your interview guide or adjust it to capture the phenomenon of interest more precisely or fully. Another advantage of the qualitative approach, as we've mentioned before, is that it can be *iterative*. Drawing from what you uncovered in your last observation or interview, you might focus your next observation or interview on a topic that needs further explanation, or you might see if a pattern you observed holds up with a different respondent or in a different social context.

For these reasons, the analysis stage of qualitative research looks very different than it does for quantitative research, which tends to stringently separate data collection and data analysis. Not surprisingly, it is much more common in quantitative research to use **secondary data**, data that other researchers collected. Quantitative researchers may spend their entire careers analyzing such data and never field an original survey of their own construction. And as we describe

further in **Chapter 14: Quantitative Data Analysis**, finding existing quantitative data to analyze is easy—government agencies and other organizations regularly track relevant trends, and sprawling online archives contain ready-to-analyze datasets. Although qualitative data archives do exist—such as the Qualitative Data Repository and the U.S. Library of Congress's American Life Histories and American Folklife Center collections of oral histories—they are much fewer in number. One reason is that, as we've discussed before, ethnographers and in-depth interviewers are part of their measurement instrument. Any data they collect is filtered through their subjective experiences, hunches, and interpretations. They determine what is true, what is connected to what, and what it all means. Qualitative analysis depends heavily on the researcher's personal knowledge of where the data came from—that is, their deep familiarity with the social context in which they observed the phenomenon of interest. While this is true for all data to some extent, the closeness that qualitative researchers have to their data means that the processes of data collection and analysis are often inseparable.

In this chapter, we'll give you a sense of how you can break down the text of your field notes or interview transcripts into useful bits of meaning that can be categorized as *codes*. As we'll demonstrate, generating these codes will allow you to see patterns in your data. It will flag surprises that may challenge or overturn the way you think about a phenomenon. Eventually, it will give you a sound basis on which to construct or refine theories. You can pursue this coding process by using specialized software or just by writing up coding memos in your word processor. We'll go over some of the basics of these different approaches. We'll end the chapter by talking about best practices for writing up your qualitative data in a clear and accessible form.

Qualitative data analysis can be thrilling because of the intense creativity that is involved. Your sociological imagination can lead you to a number of "a-ha" moments when you piece together surprising connections or come up with compelling explanations. If you're lucky, the development of such theories will inspire other scholars and launch new lines of research. Moreover, the findings of qualitative research can be particularly rewarding for research participants, who may appreciate getting a bird's-eye view of their everyday experiences—one that is rooted in the actual facts of their lives, rather than imposed from up high by theorists. With that power, of course, comes the responsibility of analyzing your data both creatively and rigorously.

Transcribing Interview Data

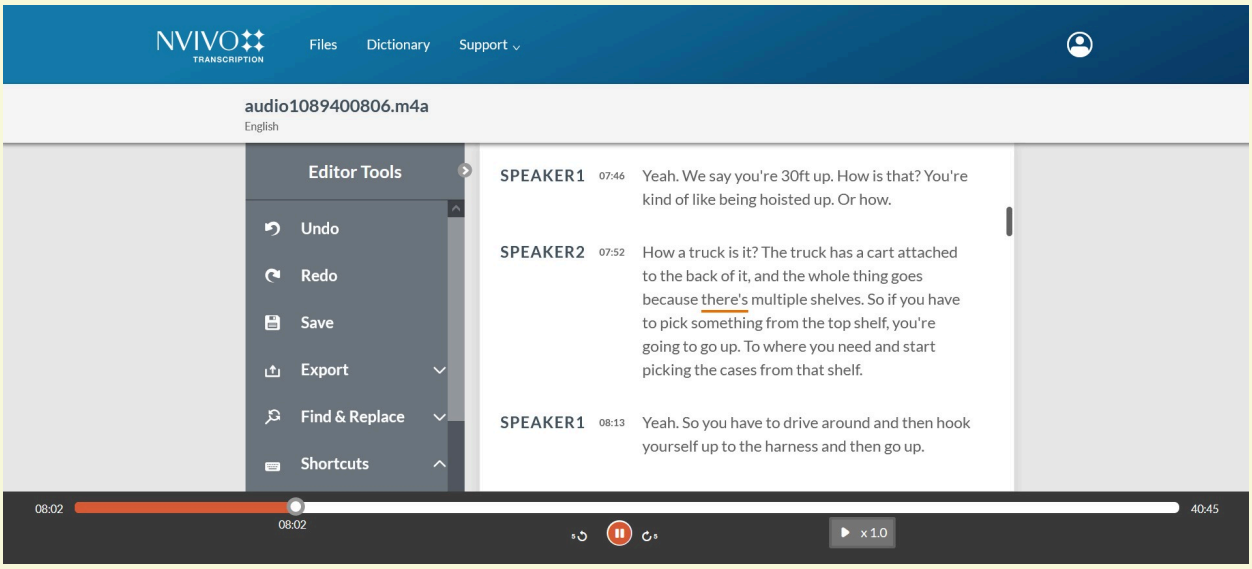


Figure 11.1. Automated Transcription Using NVivo. The qualitative data analysis program NVivo offers a paid transcription service via the cloud. Its artificial intelligence–powered algorithms can transcribe audio files—in multiple languages, and with multiple speakers—in a matter of minutes. As seen in the screenshot, the automated transcription is not perfect (in the first line "Let's say" has been transcribed as "We say," for instance). But assuming your audio is of a high-enough quality, automated transcription can substantially reduce the amount of time you need to devote to the task of converting your audio files into text. Services like NVivo Transcription also let you compare a transcript with the corresponding audio and make any necessary edits to the text, allowing the software to learn to improve its transcription accuracy. (In this example, the audio is being played back, with the orange underlined text marking the program's current location in the transcript.)

When qualitative data analysis involves interview data, the analysis usually begins with transcribing the audio recordings of those sessions. During this process, a **transcript**—a written, usually word-for-word rendition of the conversation—is created by playing the recording back and typing in each word that is spoken, noting who spoke which words. In general, it is best to have a verbatim transcript, one that reports exactly what was said in the recording. This is because the specific wording of a statement can matter greatly in how you interpret its social meaning. Furthermore, verbatim transcripts give you the option of including more direct quotations and concrete details in your write-up, allowing you to take full advantage of the greater depth of description that is possible with qualitative methods.

Many qualitative researchers choose to transcribe their audio recordings by themselves. This has a number of advantages. You will recall considerably more details from an interview you transcribe yourself. You will also be able to make a note of background behaviors and interactions that may be relevant to your analysis but are not picked up by the audio recording. As we discussed in **Chapter 10: In-Depth Interviewing**, your interviewees

might roll their eyes, wipe tears from their face, and even make obscene gestures. These nonverbal cues speak volumes about participants' feelings, and unless you write them down in your field notes or include them in your transcript, they cannot inform your analysis.

As useful as it can be, transcribing your own interviews takes a great deal of time, and it can be mind-numbingly boring. You will need to replay the recording many, many times if you are doing a verbatim transcription. Alternatively, you can hire professional services to take care of the transcription for you, but they tend to be prohibitively expensive for novice researchers. A middle ground between self-transcription and professional transcription is using the automated transcription services offered by a number of websites and applications, which convert audio into text using artificial intelligence. Automated transcription can be relatively inexpensive, and some services, such as Otter and Transcribe by Wreally, offer limited access to their services for free. As underscored by the example in **Figure 11.1**, however, you will definitely want to look over carefully any document that is auto-transcribed (and, really, any transcript you don't generate yourself) before including it in your analysis. In addition to ensuring the accuracy of the transcription, this review can provide some of the benefits of transcribing interviews by yourself—namely, refamiliarizing you with the raw data you collected and giving yourself an opportunity to insert additional details. Directly in the transcript document or in an accompanying field note, make sure to write down any observations you remember concerning your participants' tone of voice, gestures, and mannerisms, along with any other relevant nonverbal information from the interview (or from before and after it). Also describe when, where, and how your participants emphasized any of their spoken words—transcription services typically won't capture stresses or idiosyncratic inflection.

While qualitative researchers typically record their interviews, sometimes this is not possible. Perhaps the interviewee refused to allow any recording, or the interviewer felt that doing so would make the interviewee skittish or otherwise change their responses to such an extent that it might harm the veracity of the interview. As noted, you may also be worried about the time or cost involved in transcribing interviews afterward. In these cases, taking good notes can be sufficient, particularly when it is not so critical to get a verbatim quote. That said, we still strongly advise you to record interviews if at all possible—for the sake of accuracy, and to prevent yourself from having to frantically write everything down during your conversation.

Opening chapter image credit: *Suzy Hazelwood, via Pexels. Adapted by Bizhan Khodabandeh.*

11.1. Coding Qualitative Data



At its most basic level, *coding* involves breaking down your qualitative data by putting "like" bits of text into thematic buckets. You continue with this process of sorting through your data until it eventually sparks ideas that lead you to more focused theoretical arguments. *julie aagaard, via Pexels*

Learning Objectives

1. Identify different strategies for coding the text of interview transcripts and field notes.
2. Describe the approaches of open and focused coding in qualitative data analysis.

Whether the analysis process occurs in the field or after leaving it, you start by reviewing your transcripts and field notes. As you read, you try to identify codes in the text. A **code** is a label (a word or phrase) that tells us what a portion of text is saying or implying—what *theme* it conveys. Put another way, a code provides a shorthand representation of some more complex set of issues or ideas. Each code is linked to specific portions of the text (called a **coding unit**), which can range from a single phrase to multiple paragraphs. Imagine that you highlighted some text, and then labeled that highlighted text; that label—that word or phrase—is the code, and the highlighted text is the coding unit. Many different coding units can share the same code, allowing you to see how a specific theme plays out across your data.

The process of identifying codes in qualitative data is called **coding**. What's the point of coding? Ultimately, you will arrive at an outline of the specific arguments you can make based on the data you analyzed. To get to that point, you'll need to follow a process not unlike what we described in Chapter 5: Research Design for literature reviews: categorizing "like" bits of information into buckets (the "buckets" here being your codes). As you add more of your data to these buckets, you will start to see similarities and differences across the material stored in each one. This might prompt you to further divide an existing code into several codes that together describe the assorted dimensions of a phenomenon, or the multiple varieties of an idea. (The Greek root for the word "analysis" means "breaking up," which fits the spirit of this approach of decomposition and dissection.)



Qualitative data analysis (QDA) programs make the task of coding easier, though for smaller projects, a manual coding process can suffice.

There are many types of coding procedures (Corbin and Strauss 2007; Esterberg 2001), but one typical approach is to begin by doing **open coding**. This is a process aimed at *identifying* concepts or key ideas that are hidden within textual data and potentially related to the phenomenon of interest. When you do open coding, you read through each field note or transcript, line by line, and make a note of whatever categories or themes jump out to you. You may want to identify types of events or incidents, label ideas or actions, or flag perceptions or interactions.

Over the course of the coding process, you will read and reread (and reread some more) your transcripts and field notes. Familiarizing yourself with the data in this way will give you a clear idea about what themes are actually present in the evidence you've amassed. It will eventually generate a number of insightful codes that become the basis for one or more theoretical arguments. But as you might already be thinking, getting to that point—wading through all your data—will also be quite a chore. Just as quantitative researchers rely on the assistance of specialized programs to sort and analyze

their data, so too do qualitative researchers—in their case, qualitative data analysis software packages, such as ATLAS.ti, Dedoose, MaxQDA, NVivo, and QDA Miner. These are programs (sometimes called QDA *programs*, QDA *software*, and so on) specifically designed to help qualitative researchers organize, manage, and code large amounts of qualitative data. They allow you to import transcripts and other text files into a workspace. You can then easily label material, cut and paste passages, search for various words or phrases, and see wherever and how often a code was applied. More advanced users can take advantage of other features, such as the ability to code multimedia files, visualize relationships within a network of codes, create word clouds, and annotate your data in various ways.

For small qualitative projects, you don't have to use QDA software. In fact, as we'll demonstrate, you can produce an effective and fairly sophisticated coding scheme just by using a word processor like Microsoft Word. (This is known as **manual coding**.)

The sections that follow will review some fundamentals of coding. We'll provide general guidance about applying codes with the help of QDA software as well as doing it the old-fashioned way. For the sake of brevity, however, our approach will be simplified, and we will not delve into more complex strategies recommended by expert qualitative researchers. For more information, we suggest that you pick up the many detailed reference books on qualitative data analysis, including the seminal textbook by Matthew Miles, Michael Huberman, and Johnny Saldaña (2019).

Descriptive Coding

In our bare-bones approach to coding, we distinguish between two general types of codes: *descriptive* and *explanatory*. Generating **descriptive codes** (also called *index codes*) is relatively easy. You see if any relevant themes come up as you read through the data—say, topics like "class," "gender," or "race," or types of behavior or thinking like "violence," "patriotism," or "racism," or any other issues or categories related to your research question. Each theme becomes the label for a code, and you should make a note of any bits of data that fit under that code—whether it's just a sentence or entire paragraphs of text. (You can even code a single word if you'd like, though it's better to code at least a full sentence so that you have more context about your data later in the process, as we'll talk about later.)

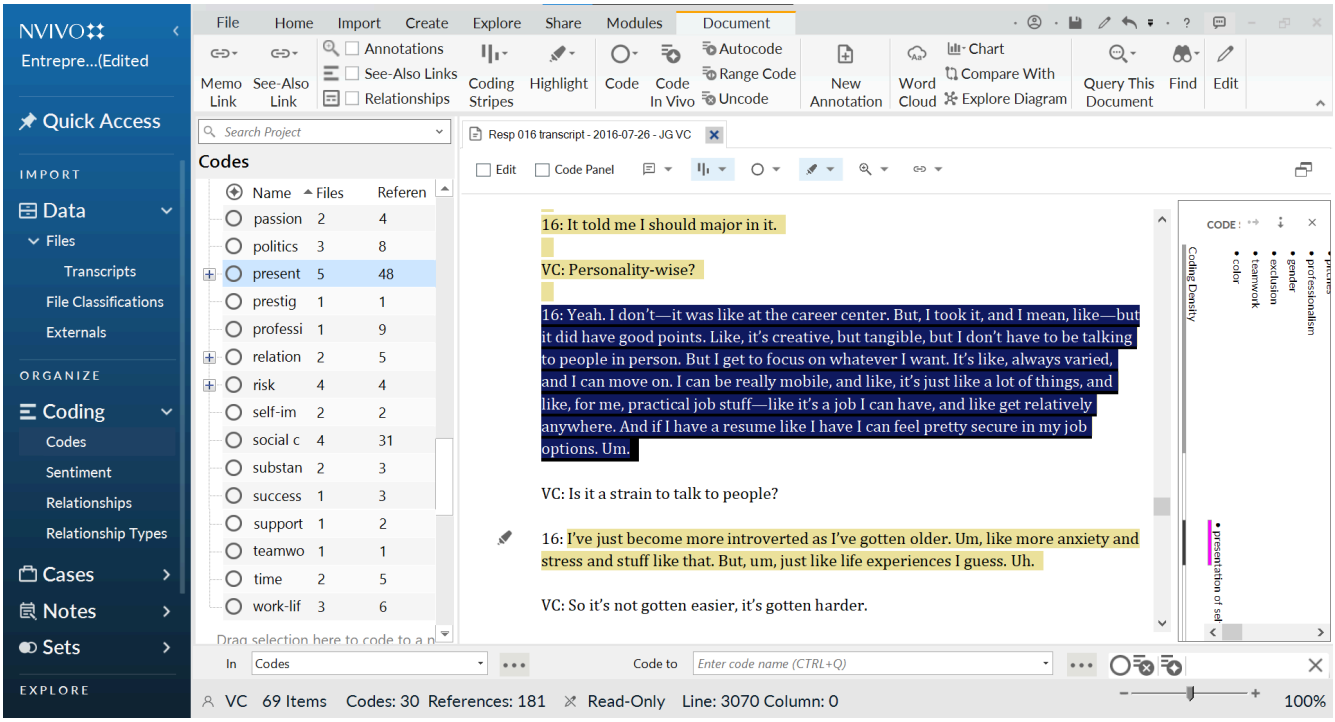


Figure 11.2. Coding Using NVivo. In the NVivo software package, coding is easy. You highlight some text (as shown with the blue text block) and then click on the "Code" button at top. NVivo will then prompt you to assign a code (label) to that coding unit (the highlighted text). You can create a new code or choose one of your existing codes (listed in the menu to the left); you can also assign multiple codes to the same coding unit. NVivo's display settings can be set to highlight any text you have coded in yellow, with the specific code identified on the corresponding stripe to the right (in the example, note the pink *presentation of self* stripe).

So how do you go about coding specific passages? If you are using QDA software, the process is simple: select the text and then click on the "Code" button (or whatever the equivalent is for your program). You can see what this looks like in **Figure 11.2**, which features the QDA package NVivo. Once you select the text and attach a code to it (or more than one code, if you wish), the program will label the text accordingly. You can code other bits of text with the same code. Then, when you're finished with your coding, you will be able to generate a document that contains—in one place—all the text associated with your code (see **Figure 11.3** for an example). This makes it easier to see any patterns in the data and to copy and paste relevant quotes or other data into your actual paper.

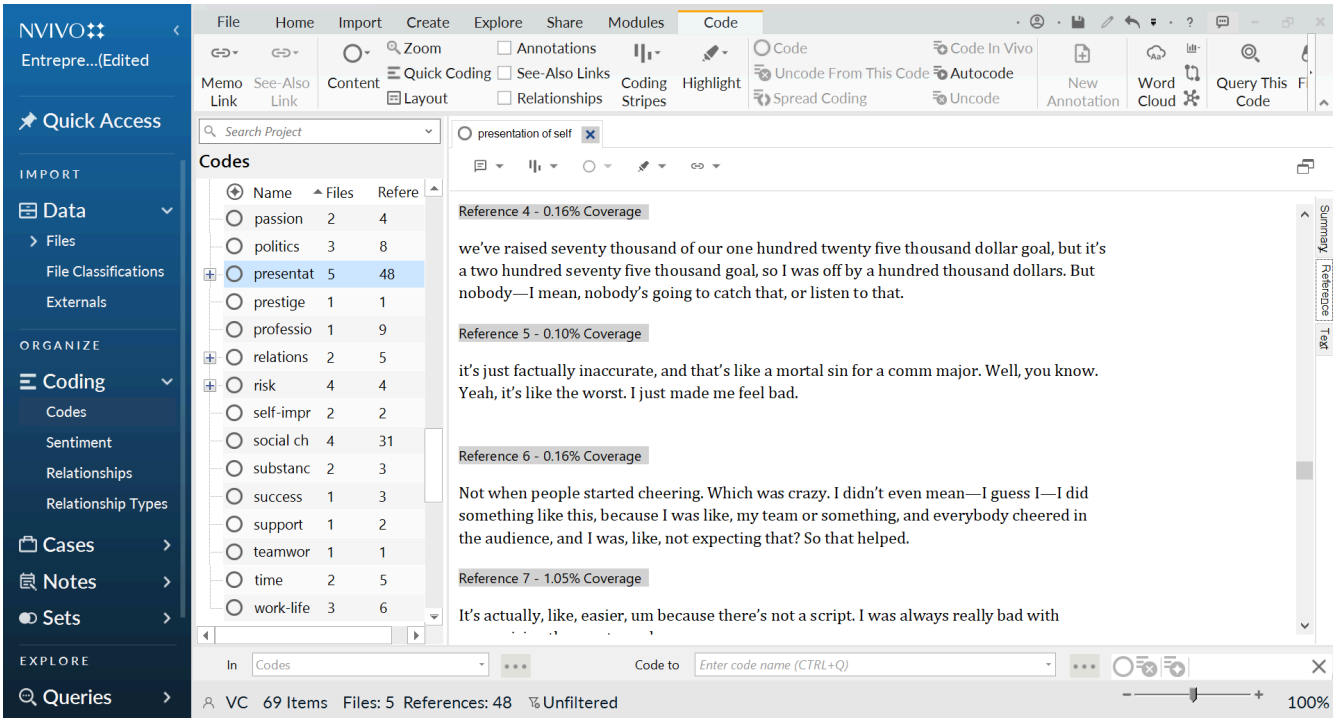


Figure 11.3. Compiling Coding References. QDA programs can automatically pull all the text labeled with a particular code (in this example, *presentation of self*). The "references" in the screenshot are the specific coding units, the bits of text that have been labeled with that code (which may come from multiple interview transcripts). You can export the document of coding references to Microsoft Word or another program and then use it as the basis for a *coding memo*.

You can follow a similar coding process using a word processor like Microsoft Word. For instance, you can create separate documents for each of the major codes you plan to use, and then you can copy and paste the associated coding units into each document. (Of course, you'll probably want to read through a few of your transcripts first and generate a list of pertinent codes before you start copying and pasting anything.)

How do you decide what specific codes to use? Start with broad topics that have something to do with your research question. If you are studying how social class affects people's experiences in high school, for instance, you might label any passages referring to those experiences with the—very broad—code *class*. (We'll talk about how to sharpen the focus of your codes shortly, but for now, it's fine to be somewhat broad.) You can also draw on any relevant theories you are aware of—say, using the code *presentation of self* when you spot instances of people fretting over their public personas, or the code *emotional labor* when respondents talk about managing their feelings at work. Another way to generate codes is to let them emerge from the terms or phrases that your research participants themselves use (called **in vivo codes**). These codes may reflect particular notions or beliefs that are common among the groups you are studying (what are called **folk theories**).

The words or phrases you use for your codes are completely up to you. You are free to create your own naming conventions, and as we'll see, it is actually a good thing when you depart from the literature and brainstorm your own codes—your own terms and theories. That said, researchers often use codes taken directly from the literature they're immersed in to ensure that they are capturing all the aspects of the phenomenon of interest already known from past studies. Such standardization can make the process of coding seem less arbitrary to outsiders. Regardless of which approach you take, it is important to identify what criteria you are using for each code. (You may even want to write this

down in a **codebook**, a list of instructions you use for coding text, which is especially useful if you are working with other people.) Being clear about the criteria for each code will not only add rigor to your coding, but it will make it easier to see if similar concepts can be grouped together (a topic to be discussed later).

As you can tell from the examples of codes we've provided so far, some may be simple and unambiguous, while others may be complex and viewed differently by different participants. For instance, whether a passage is about "work" may be pretty straightforward, but whether it reflects the concept of "emotional labor" may be more of a judgment call that different coders may disagree about. The sizes of coding units also vary widely. Simple codes, such as "organizational size," may refer to just a few words of text (a quote to the effect of, "Our company has 500 employees"), while complex codes, such as "organizational mission," may be assigned to a long text block of several paragraphs (perhaps encompassing the entire answer to a question you asked of a nonprofit director about their group's mission).

As you proceed with your coding, it is important that you not let your original research question or tentative hypotheses cloud your ability to see categories or themes. You should actively seek out new concepts relevant to the phenomenon of interest. It's called *open* coding for a reason—keep an open mind. You may have even noted some preliminary ideas for codes in your field notes or journal entries. Take full advantage of these hunches. Once a basic set of codes is identified, they can then be used to code the remainder of the data, while simultaneously looking for new concepts and refining old concepts.

A specialized type of descriptive coding is called **attribute coding**. In this approach, you flag important characteristics of the case that was observed. When analyzing data from an interview-based study, for example, you likely will want to capture demographic information about your respondents, such as their gender, age, or race. If you're using a QDA program, you can add codes to indicate where in each interview transcript that information is conveyed. If you are coding manually, you can simply record the details in a separate spreadsheet that lists all your interviewees and their demographic characteristics, as shown in **Figure 11.4**. (Even if you use a QDA program, you still might want to create such a spreadsheet to ensure that these key details are easily at your fingertips.)

Age	Gender	Race	Ethnicity	Immigrant?
21	female	white		
19	male	white		
24 (calculated)	male	South Asian	Pakistani	1st generation
18	male	South Asian	Pakistani	
20	female	white		
21	male	white		
NF	male	white		
20	male	African American		
NF (graduated HS	male	white		
26	male	African American		
29	male	Middle Eastern	Afghan	Yes
22	female	African American	Ghanaian	2nd
NF	male	African American		
22	female	white		
22	male	East Asian	?	
NF	male	white		
NF	male	white	italian-american	2nd gen
21	female	African American		

Figure 11.4. Interviewee Spreadsheet. During the coding process for an interview-based study, you will want to capture demographic information and other relevant individual characteristics relating to your sample of interviewees. You can store these details in a spreadsheet, in addition to coding the specific answers to such questions in your interview transcripts (through a process called *attribute coding*).

What's the point of attribute coding? For one thing, the demographic data you are organizing through this process will be crucial when you describe your interview sample in the methods section of your paper. Readers will want to know how old your respondents were, how many were white and nonwhite, and so on. Also, you can use these attribute codes to help you discern patterns in your data. For instance, do you see a tendency for people of a certain sexual orientation, gender, race, or class to answer one of your questions in a certain way? If so, that could be an interesting finding to explore, and your coding process will have led you to that insight. (At the same time, remember that you never have a truly representative sample in a qualitative study, so any patterns among your respondents should be considered only suggestive and should not be presumed to hold up more broadly.)

We should reiterate at this point that qualitative analysis often happens when the researcher is still collecting data—sometimes, right after every interview or observation session that is conducted. For this reason, don't think of

the process of coding as something that has to happen after you leave the field. In fact, as we've suggested, there are major benefits to quickly reviewing your data as soon as you can, such as being able to adjust your approach for future observations and interviews based on what you already know. For any such preliminary analysis, you may not have access to full transcripts, and you won't yet have an exhaustive list of codes at hand to draw from. Maybe you'll just be able to jot down a few analytical points or questions that you wind up including in your field note. That's fine. Try your best to integrate the techniques we have described whenever and wherever they are feasible.

So far, we've talked about descriptive coding—describing bits of data as reflecting this or that category, idea, or theme. As we discussed in earlier chapters, however, sociologists usually want to go beyond just describing reality. They want to explain it. Specifically, they want to make theoretical arguments that explain why they see what they see in the social world, or why the phenomenon they're interested in unfold in the ways they observe. This goal of explanation means we'll ultimately want to kick our coding up a notch and devise more ambitious and insightful codes. That said, when you begin the process of open coding, you should expect to be doing descriptive coding for the most part—categorizing the data rather broadly, with a focus on index and attribute codes. By coding in this way, you will become intimately familiar with the data. Eventually, inspiration will strike. You will think up ideas for actual arguments, and your coding will start moving in that direction. It may take several readings of your data to get to this point, however.

Explanatory Coding





The coding process entails breaking down the text you're analyzing into bits of information that can be categorized thematically. It can also involve connecting the codes you create—merging and organizing them to build overarching theories. *Alexa Popovich, via Pexels / Zoltan Kluger, via Wikimedia*

To be honest, there isn't a clear line between descriptive codes and **explanatory codes** (which are sometimes called *analytic codes*). You can think of this distinction as more of a continuum, with description on one end and explanation on the other. You gradually want your codes to move in the direction of explanation, which means not just categorizing data but also making arguments based on it. For instance, you should think about how or why two or more variables that you've uncovered in your coding are linked. You should also be on the lookout for new themes and ideas that may differ from, or complicate, your original research question or theory. As we've discussed before, this is a key advantage

of qualitative research: its ability to surprise the researcher. With these goals in mind, here are some questions you can ask about each bit of text you review:[1]

1. What is this a case of? What topic or aspect of the phenomenon you're studying is it an example of?
2. Does it raise any questions about the phenomenon you're studying? Does it answer any questions (e.g., providing an explanation)?
3. Does it contradict or complicate the existing theories? (Remember our discussion in **Chapter 4: Research Questions** of the **abductive approach** to analysis—looking for any "surprises" in the data, given existing theories.)
4. Does it speak to a new theory? (As we will discuss later, don't be afraid of brainstorming *neologisms*—your own way of succinctly describing a phenomenon—or refining existing terms.)

1. This list builds upon an earlier set of suggestions by John Lofland and his coauthors (2005).

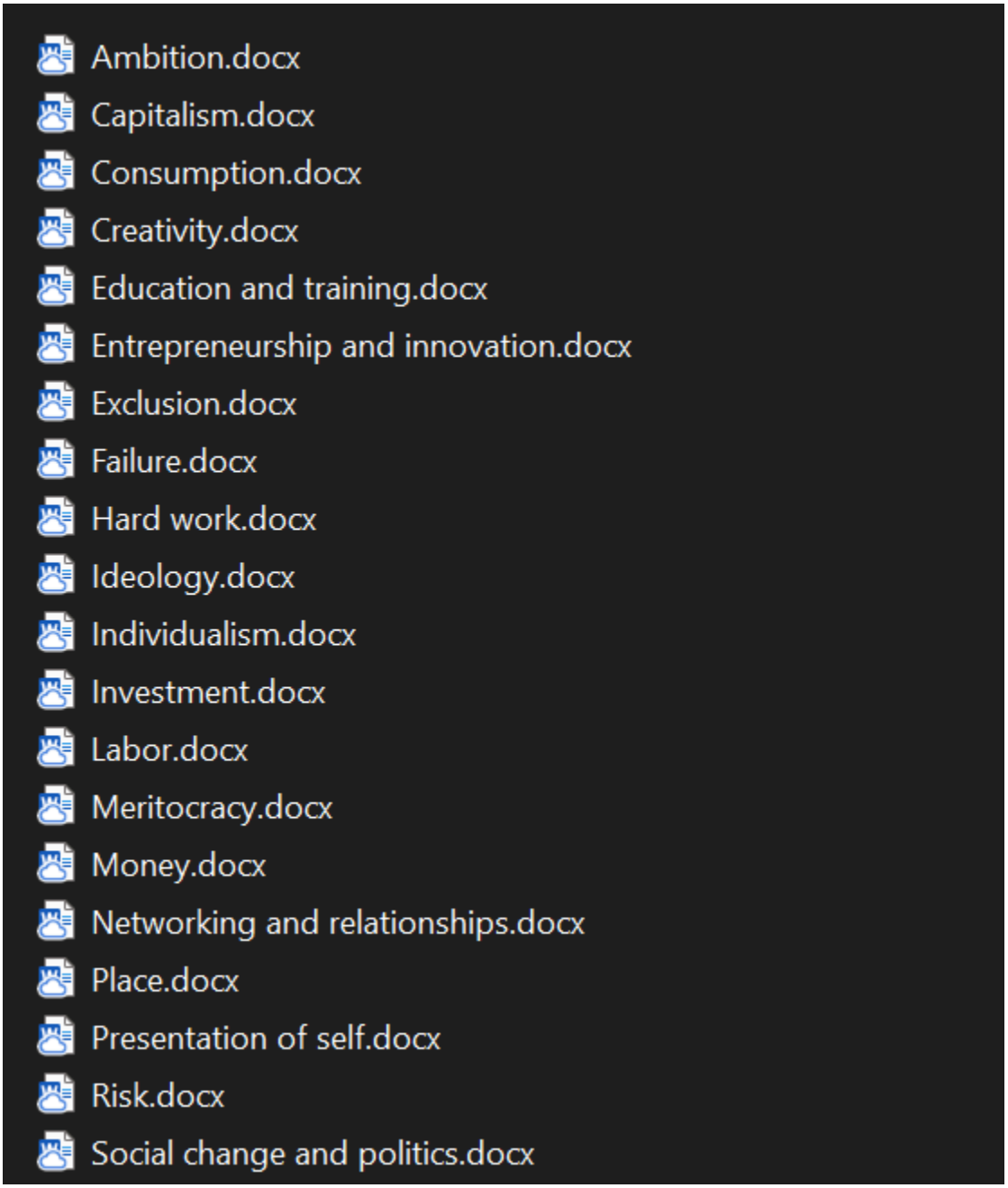


Figure 11.5. Folder of Coding Memos. Whether you are coding manually or using a QDA program, you should draft *coding memos*, documents that explain what your codes mean and relate those ideas to the actual data you've gathered. In a word processor, you can create documents whose filenames reflect your key codes (as in the folder of files displayed in this example). These memos can contain some or all of the text you labeled with that code (compiled by your QDA software or manually copied and pasted into each document) as well as your preliminary thoughts about any patterns or relationships in the data.

Through the process of opening coding, you will eventually wind up with many separate documents full of relevant data, each focused on a specific code. (In the QDA workspace, these would be the documents the program compiles of all the text associated with each code.) You can use these documents as the basis for **coding memos** (also called *analytic*

memos). Coding memos present your preliminary ideas and hunches about the data you're reviewing. For instance, let's say you are coding manually and create a number of Word documents titled with the words or phrases of your most important codes, such as *presentation of self* (**Figure 11.5** illustrates how these files might be organized on your computer). These documents are the broadest categorization of your data.

Now let's dive a bit deeper. In the *presentation of self* document, suppose you have copied and pasted all the data from your interviews and field notes relating to this broad theory. You start going through the material in the document, noticing connections between your respondents' quotes here and there. As you read, you organize the material so that "like" bits of information are together, creating sections within the document that get at more specific dimensions or understandings of the presentation of self across your data. As you continue to categorize your data in this way, your thinking about the topics in your codes becomes more tightly focused. As a result, you start to create subsections within the sections of the document, breaking down each concept into narrower ones—or even, if you're lucky, into specific arguments about what patterns exist in the data or what factors might explain those patterns.

Through the sort of process we've just been describing, the sections (and subsections and subsubsections) in your coding memo will start to take the form of explanatory codes. For instance, let's say you create a section within your *presentation of self* memo devoted to the topic of authenticity, which gets at the tensions that people feel about being "fake" in their public personas (see **Figure 11.6** for an illustration). Then you start looking at all the data you've categorized in this authenticity section and realize that there are several dimensions of authenticity to consider: authenticity as consistency (not being erratic, but showing a single persona); authenticity as sincerity (showing your "true" self); authenticity as passion (showing genuine excitement about the work you're doing); and authenticity as intimacy (revealing a raw and unfiltered self). These become subsections in your *authenticity* section, and at this point, you might be wondering which of these dimensions of authenticity are more relevant to the phenomenon you're studying, or which dimensions show up in particular social contexts. Now you're generating tentative theories.

As you further refine your codes, you may tweak your definitions of them (e.g., how exactly should you define "authenticity"?) and what types of data would fall into each one. Pay special attention to any **outliers** or **deviant cases**—instances in which a particular case (e.g., person, organization, observation) doesn't fall into the existing theory that you or other scholars have devised to explain a relevant phenomenon. The sociologist Howard Becker (1998:87) aptly summed up this general approach: "identify the case that is likely to upset thinking and look for it." In qualitative research, these instructive outliers are incredibly helpful because they show you where existing theories fall short—and, therefore, where they can be supplanted by new theories. Remember as well the point made in **Chapter 4: Research Questions** that there are no "negative" cases, only positive cases waiting to be explained.

Entrepreneurial sensibility as inauthentic

The popularity of the "personal brand" approach to identity work has a pernicious effect in that it limits the creation of self to a commercial self, rather than sense of individuals as unique human beings with manifold (non-market) skills and personal traits. It also requires "practicing" one's pitch (i.e., the Personal Branding Statement, or PBS) all the time; the PBS approach is similar to entrepreneurship, in that founders talk about the all-consuming nature of their work: how they could be doing stuff related to their business at any moment, and therefore need to really work at a work-life balance.[7]

R71 talks about how she needs to wear makeup and professional clothing and otherwise work on her physical appearance in order to take on her role in her startup. "I look Photoshopped," she says. [8] It is interesting because her usual look—how she dresses for class or for meetings with us—is so low-key and college kid-like. She speaks in an excitable manner, sometimes with a degree of (self-admitted) brassiness and spunk, repeating herself for emphasis at times, and yet she comes across as a bit self-critical during moments when she berates herself, with a smile or laugh, for mistakes she's made (she admits she has some self-esteem issues). On that day, she wasn't wearing much, if any, makeup, in contrast to when we saw her the week before, when the foundation was noticeable at the wrinkled inner corners of her eyes as a bluish sheen.[9]

R71 points out that she wasn't being "dishonest." But someone told her once, "you need to just go from, like zero to a hundred when you have to."[10] The management style that they promote is to "overpromise and underdeliver."[11] After pitching at a tech event, she says she didn't feel she was living up to her "costume"—here she means the physical appearance of dress and gestures that she "put on" when she went to give the pitch. She

Figure 11.6. Coding Memo. This coding memo devoted to the code *presentation of self* has been divided into thematic sections. The section shown here discusses how taking on an entrepreneurial persona of hustling and self-promotion can make student entrepreneurs feel "inauthentic." (Note that the details in the anonymized excerpt have been slightly altered to preserve confidentiality.) Breaking down a broad descriptive code into more focused codes can give you a richer sense of the different dimensions of a phenomenon and help you start making pointed arguments about your data.

By refining your codes, you are moving away from description and toward explanation. So far, we have talked about slicing up broad descriptive codes by describing different dimensions of each code, which is in line with the idea of "analysis" as "breaking up" data through categorization and labeling. Note, however, that you can also merge codes: creating higher-order codes that link multiple codes and thereby create larger "buckets" of information. (In grounded theory, the lower-order codes are called *concepts*, while the higher-order codes are called *categories*.) For instance, you might connect ideas about "authenticity" in your *presentation of self* code to a separate idea about *personal branding* (i.e, the notion of marketing one's professional or personal identity to specific audiences). Rather than further breaking down your data, here you are drawing higher-level connections between different ideas (perhaps even visualizing these connections in a **concept map**, as described later). By thinking up these higher-order codes, you helpfully reduce the number of ideas you are working with and paint a "big picture" of the key factors necessary to understand a particular phenomenon.

Just as with breaking down codes, creating overarching categories can be done in phases—for instance, by combining codes into a set of higher-level codes, and then merging them into still higher-level codes. Note that QDA software makes the process of both merging and splitting codes very easy. In most programs, for instance, you can nest one code within another by simply clicking and dragging within the list of your codes.

when we walked in the room, they greeted us, they followed up immediately after. . . . We think their idea is pretty crappy, but that's the type of initiative we want to award, so we're going to give them 2,500 bucks.'" These judges rewarded the team's eager advances despite the absence of a compelling project. The lesson, then, is that entrepreneurial potential is not simply a matter of innovative vision, but perhaps even more a function of one's ability to nurture relationships with investors.

In the theater of entrepreneurship, the whole world is potentially an investor—even though, of course, most people are not. And the successful entrepreneur is always in character—always pitching, always presenting themselves as investable. The anxiety of constantly having to be "on" for this audience was very real to our student entrepreneurs (Marwick 2013). Kayla, a 20-year-old white VCU student from a (as she put it) "maybe lower-middle-class" background, talked with ambivalence about the work she put into her physical appearance as the director of business development for her startup. Her usual look was low-key and typical of a college student, but whenever she had an entrepreneurship-related event, she would wear heavy makeup, put on professional clothing, and refine her look to the point that she looked "Photoshopped," as she put it. Moreover, she would take on the persona of an entrepreneur—or what she perceived to be that persona—by self-consciously acting more "adult" in the words she chose and mannerisms she used, toning down her typically explicit language and penchant for silly jokes. She described this overall package as her "costume"—the dress and gestures that she "put on" whenever she operated in entrepreneurial circles, in order to feel more professional and "powerful." Going through the ritual of adopting that persona gave her the confidence she needed to perform well in these spaces, she said. Of course, a focus on crafting one's appearance to please others is nothing out of the ordinary for workers (or women more broadly), and as we explore in other work, there is a heavily gendered dimension of the presentation of self in entrepreneurial circles. In any case, Kayla saw the finessing of her appearance and mannerisms as part of a larger obligation to perform for investors. She brought up how a New York entrepreneur who spoke at an event put on by VCU's startup club said that founders needed to learn to "go from zero to a hundred" whenever they were in spaces with potential investors: they had to be ready at all times to sell themselves.

Figure 11.7. Published Manuscript Incorporating Coding. In this excerpt from the article published in *Social Problems* (Chen and Goldstein 2024), you can see how some of the material included in the coding memo in **Figure 11.6** wound up being used in a more polished form. *Oxford Academic*

What is the payoff for reorganizing and manipulating codes in all these ways? You will end up with a number of higher-level codes with lower-level codes nested within them. Some of these categories will speak to larger arguments about how or why the variables in your analysis might be connected—that is, what causal relationships exist, and what causal mechanisms operate within those relationships. As you keep nesting subarguments within your overall argument, you will start to flesh out your empirical analysis, generating key takeaways and supporting points for the paper that eventually emerges. (**Figure 11.7** shows how a paper can build upon the preliminary analysis in coding memos.)

Beyond this general advice, there are three analytical strategies you can take to generate theoretical arguments: neologisms, typologies, and concept mapping. We'll discuss those approaches later, but first let's wrap up the discussion of coding by talking about the second stage: focused coding.

Deeper Dive: Axial and Selective Coding

In addition to open coding, grounded theory experts Juliet Corbin and Anselm Strauss (2007) recommend two other coding techniques: *axial coding* and *selective coding*. You can apply all three procedures at the same time. When you conduct **axial coding**, you assemble categories and subcategories into causal relationships or hypotheses that can tentatively explain your phenomenon of interest. (Remember that in grounded theory, lower-order codes are called *concepts*, and higher-order codes are called *categories*.) The relationships between these categories may be clearly evident in your data, or they may be more subtle and implicit. To clarify them, you can differentiate between categories that represent *conditions* (the circumstances that generate or sustain the phenomenon), *actions/interactions* (the responses of individuals to events under these conditions), and *consequences* (the outcomes of actions/interactions). As you identify conditions, actions/interactions, and consequences across your data, you will start to generate overarching theories to explain why a phenomenon occurs, under what conditions, and with what consequences.

Another coding technique is **selective coding**, which involves identifying a central category or a core variable and systematically relating it to all your other categories. The central category can evolve from existing categories or can be a higher-order category that subsumes previously coded categories. New data is selectively sampled to validate the central category and its relationships to other categories (i.e., the tentative theory). By utilizing selective coding, you helpfully limit your analysis, prioritizing what's truly important. At the same time, you need to keep an eye out for other categories that may emerge from the new data, which can signal that your initial theory needs to be further refined.

Moving from Open Coding to Focused Coding

Open coding will probably require multiple rounds of reading through your transcripts or field notes. The process, as we've described, is about inductively identifying patterns in your data, and refining and redefining your codes as you go. At some point, however, you will want to begin **focused coding**, a process of firming up and prioritizing the codes you developed during open coding and applying this finalized framework to the entirety of your data. This shift typically occurs when you reach **saturation**—that is, when additional data no longer leads you to revise the ways you perceive core categories or relationships.

Focused coding is a multistage process. First, you should read through the coding memos and other notes you made during the open coding process. Use them to help you collapse, or narrow down, the codes you identified across your data. Identify themes or categories that seem to be related, and see if you can merge some of them. Next, settle on your priority codes—the ones that seem most promising, given your research question and theoretical orientation toward the phenomenon being studied. Once you arrive at a short list of codes, make sure each one has a definition that clearly spells out what the code means, perhaps including these definitions in the coding memos you've already started or in a separate codebook. This process of explicitly defining codes will not only help you make further sense of your data, but it will also give you more confidence when talking about your findings. Finally, recode the dataset using your short list

of codes, making sure that you apply the definition of each code consistently throughout each transcript or field note you are recoding.

In open coding, you are constantly creating new codes and tweaking the definitions of existing ones, but once focused coding begins, the definitions should not change. Any deviation will make the data analysis less trustworthy. That said, if there are pieces of data that do not fit with your revised categorization scheme, then it is important to note those deviant cases—either listing them as exceptions in your final report, or going back to the open coding stage and rethinking the theories you generated.

Especially at the focused coding stage, using multiple researchers to code your qualitative data can be quite helpful. You may miss something a participant said that another coder catches. Similarly, you may shift your understanding of what a code means and not realize it until another coder asks you about it. Having multiple coders also gives your analysis greater credibility, because readers will have less reason to believe you cherry-picked your data to support your preferred arguments. With more coders come more logistical problems, of course—the research team must come to a consensus about the meaning of each code, and they also have to find a way of ensuring that everyone is applying the codes consistently across the data. (We discussed these sorts of issues previously when we talked about **inter-rater reliability**.)

When reading our descriptions of open and focused coding, you might have been struck by how labor-intensive these procedures are. We should point out that we have outlined best practices. To be honest, ethnographers do not always have the luxury of time and resources to pursue by-the-textbook analysis. Fieldwork is demanding in itself. Furthermore, qualitative data analysis can be exciting for the novel theoretical discoveries and rich empirical detail it generates, and you don't want to stamp out any joy in that process by imposing an overly fastidious and pedantic approach. That said, it's important to be as rigorous in your analysis as you can be so that you can ultimately be confident that your theories reflect the truth of your data. To that end, try your best to do preliminary analysis while in the field to ease the burden of the analytical work later on. And be patient through the long process of observation and coding. Sometimes, how a case you observe in the field relates to your larger phenomenon may not be apparent for some time. Let the data seep deeply into your brain, and eventually compelling insights will arise.

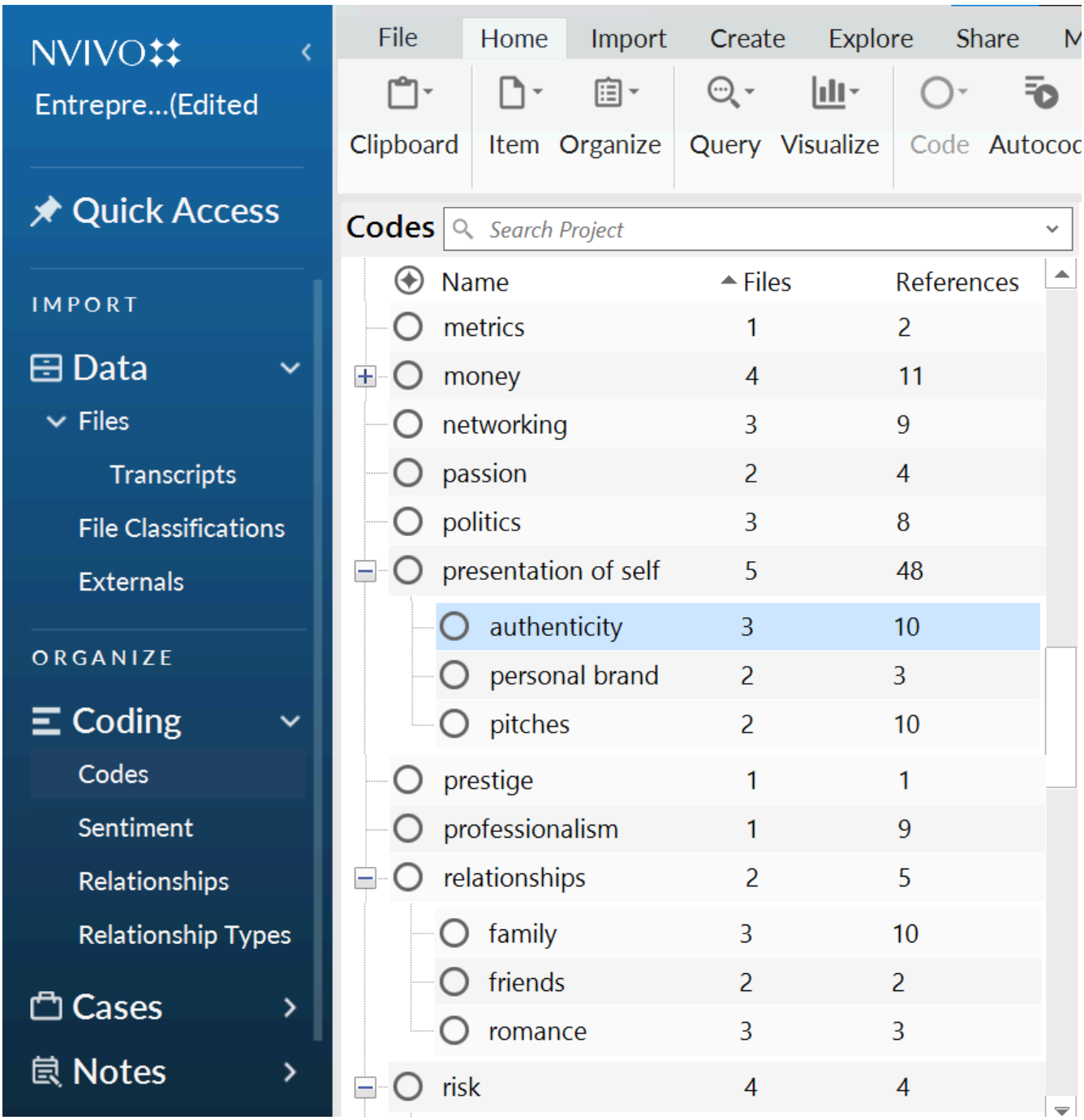


Figure 11.8. Nesting Codes. Using QDA software, you can nest one code within another one by simply clicking and dragging. You can also merge and reorganize your codes easily. This flexibility is a key advantage of using QDA programs. However, you can mimic the approach manually by creating coding memos in your word processor, each focused on a key code. Additional codes can be nested within this code as the document's sections and subsections.

A final word about open and focused coding: QDA software makes these two processes much, much easier. You can easily merge or divide up your codes, as shown in **Figure 11.8**. You can remove coding units from categories in which they no longer fit. And you can finetune your codes and apply them to large amounts of new data without much problem.

Collaborative coding, too, can be managed more efficiently, and software features can even help you evaluate the inter-rater reliability of your coding versus those of your collaborators.

QDA software can also conduct automated analyses, processing large volumes of text data using user-defined rules. (This is especially useful for qualitative content analysis, described in **Chapter 15: Materials-Based Methods**.) To automate the coding process, however, you will need to create a *coding schema* that specifies the keywords or codes to search for in the text. You'll also have to test that schema beforehand using a sample of texts to ensure it is not biased or incorrect. Otherwise, the resulting analysis may wind up being useless. Keep in mind, too, that software programs can miss sarcasm, metaphors, or other subtle meanings and contexts, which may lead to significant misinterpretations in large-scale automated analyses.

Actionable Research: A Q&A with Leslie Martin



University of Mary Washington

Leslie Martin is an associate professor of sociology at the University of Mary Washington in Fredericksburg, Virginia. A scholar of urban sociology and social inequality, Martin's research focuses on issues of neighborhood and community change. She has written extensively about affordable housing in growing cities and the social consequences of government policies that promote homeownership, drawing wide media attention for her

work. She has also studied policies and programs to deal with homelessness and the perceptions that chronically homeless people have about their safety and access to public spaces. As the faculty director of Mary Washington's Center for Community Engagement, launched in 2019, Martin has sought to strengthen relationships between the university and local organizations. The courses she teaches often incorporate hands-on, community-based learning, giving students a chance to volunteer locally while applying sociological perspectives to the analysis of community problems and opportunities. Martin received her doctorate from Emory University and a master's degree in urban studies from Temple University. Before becoming a professor, she worked as a housing counselor for a fair housing organization and an AIDS service organization. She currently serves on the board of Fredericksburg Continuum of Care, a regional coalition of homeless service organizations.

How did growing up in Richmond and attending Richmond city schools shape your sociological interests?

When I was growing up, race was always evident and on the table as an issue to be explored. As a student in the 1970s and 1980s, I was "clustered" with a handful of other white students in elementary school classrooms—an attention to my comfort not often afforded to black students in integrating schools. My middle school was within walking distance of my house, but it was also attended by students from the other side of the James River who were bussed in to achieve integration. My high school was an explicitly racially mixed alternative school. I worked on the school newspaper, and we devoted significant time to interviewing students about their views on race. All of this contributed to my belief that talking about difference and inequality was not off-limits. This primed me for the sociological endeavor.

I pursued urban studies in a master's program just because I fell in love with cities, but later I turned to urban sociology. I appreciated the explanatory tools that sociology in particular provides to help us understand urban patterns and urban inequalities.

You have conducted various types of research at the level of urban communities. What do we as sociologists gain from studying cities at the neighborhood level?

I often tell my students that they may think local issues are "boring" compared to national issues, which are more likely to make headlines, but the local is where the rubber meets the road. It's where structural inequalities get acted out and directly impact people's lives. We may be concerned about a punitive turn nationally in our approaches to homelessness, but what exactly are the consequences of the local policies now in place for people who lack housing? How does our city or town's police department actually enforce quality-of-life ordinances?

Serious analysis at the local level—and often even the neighborhood level—allows us to see clearly who has power and how it is manifested. The local level is also a place where we can see ourselves as more powerful, more influential. Working together, we can often enact change in a local community to an extent that would be much more difficult to achieve at a state or national level.

In an academic article based on your study of gentrification in Atlanta (Martin 2008), you found that some residents voiced fears about threats to their children as a socially acceptable way of objecting to the class and race backgrounds of their neighbors. How did this nuanced argument about what was left unsaid in the in-depth interviews emerge from the process of qualitative data collection and coding?

What was amazing to me was that I did not ask my respondents about children at all. I had a semi-structured interview instrument that focused on how respondents experienced their neighborhood, its changes, and their engagement with neighborhood organizations. But while I was analyzing my interview data, themes about children kept emerging—across neighborhoods, and across categories of respondents. Obviously, asking people about where they live inspires answers related to family. What I found in my data was more than that, though.

Respondents were not simply speaking about their own children. They were speaking about abstract "children" they wanted to protect.

In your current role as the faculty director of the University of Mary Washington's Center for Community Engagement, what sort of work do you do to bring together faculty, students, and community actors?

Our center is fairly new, but we are building on 30 years of student-run volunteerism at Mary Washington that has led to long-standing relationships in the community. What we are hoping to do is to make that web of relationships stronger, richer, and more complex. We have been working to become a central access point for community members and to connect people with the resources, offices, and expertise they need in new and creative ways. For example, our collaboration with Fredericksburg City Public Schools started with a student-run gift box drive during the December holidays, but we have extended that partnership with various campaigns. Early in the pandemic, we sent our students to the school district's "internet cafes," which filled gaps in the community's online access. We've recruited faculty and staff to serve community-based school lunches. And we've served as a host site for high school students with disabilities who want to gain work experience through internships.

What sorts of opportunities for community engagement do you offer students in your courses?

In some of my courses, like a first-year seminar on housing and homelessness and an upper-level sociology course on social welfare programs, students complete a set number of hours working with a community-based organization of their choosing. They then reflect on their experience and how it connects with course materials several times during the semester. For example, in the social welfare course, we end the semester by discussing their community work after they've had a chance to read "To Hell with Good Intentions" (1968), a classic address by the Catholic priest and social critic Ivan Illich, along with more contemporary writings on the charitable mindset that many people take into volunteer work. We ask how these perspectives and their own experiences relate to what we have learned about the U.S. welfare state and theories of welfare state development. The combination of their community work and the critical reflection they do in class helps them see nuance in course concepts they may not have otherwise seen.

I also teach a public sociology course where students work with community partners on research projects. The research is community-driven, which means students learn to recognize the expertise that resides outside of the classroom. They have to figure out how to adapt their academic skills to serve the needs of the community. This is challenging work for all involved, but it has proven to be important for the students and useful for our partners. For example, one student research team surveyed our university's students about their experiences of food insecurity, and the results of that study spurred the creation of the UMW Eagle Resource Closet, our on-campus food and toiletry pantry. During the first year of the pandemic, another student team partnered with a local homeless service organization to conduct interviews with people who had been temporarily sheltered in hotels. Students discovered that while the residents saw some advantages to living in hotels, they had significant concerns as well. Many had more difficulty accessing public transportation from the hotel location, felt isolated in a room alone, or wished for more contact with service providers. These findings helped the partner organization decide to pursue a different long-term strategy for housing the chronically homeless.

How would you define the kind of publicly engaged work you do, and what are the basic principles that you follow as a researcher who works with communities?

I characterize the work that I do as public sociology. I am not convinced it is always successful—this is a work in progress! I have been trying to take Mitchell's (2008) critical service-learning model to heart, by focusing on developing authentic relationships, maintaining a social change orientation, and working to redistribute power. This is a tall order, and I try my best to approach these goals in my work.

What would you personally say is the best way that sociologists can contribute to social change? How should sociologists view that goal in relation to their goal of advancing scientific knowledge?

What a tension: I'm having my public sociology students read about that very debate right now. Personally, I am not very motivated by advancing knowledge for its own sake. I am very motivated by actionable research. My conception of "actionable" can include theoretical work as well as empirical work—anything that helps us understand systems of inequality, and efforts to disrupt those systems. To that end, doing work that holds meaning for us and those around us is important, and giving thought to where and how we share our findings is also key.

Do you have any advice for students interested in pursuing the kind of sociological work you do?

Anyone can be a public sociologist, and public sociology can be done anywhere. Following your passion is the most important thing. I've always been fascinated by housing issues, and I've been able to stay engaged in housing work since my undergraduate career. On this and other topics, you'll find that there is plenty of room to do publicly engaged work—in social work or in advocacy, as a researcher or as a teacher.

Key Takeaways

1. Open coding is an inductive process of allowing codes to emerge from the data and refining them as you go along. Focused coding draws from the codes created during open coding and applyies them consistently across all the data being analyzed.
2. During open coding, you should start with broad descriptive codes but gradually move in the direction of more focused explanatory codes. The latter pose specific arguments that have been inspired by the categorization of the data you have done throughout the coding process.
3. Coding can be done manually in a word processor or by using specialized QDA software like ATLAS.ti, Dedoose, MaxQDA, NVivo, and QDA Miner, which make it easier to organize, sort, and analyze qualitative data.

11.2. Creating Arguments out of Codes

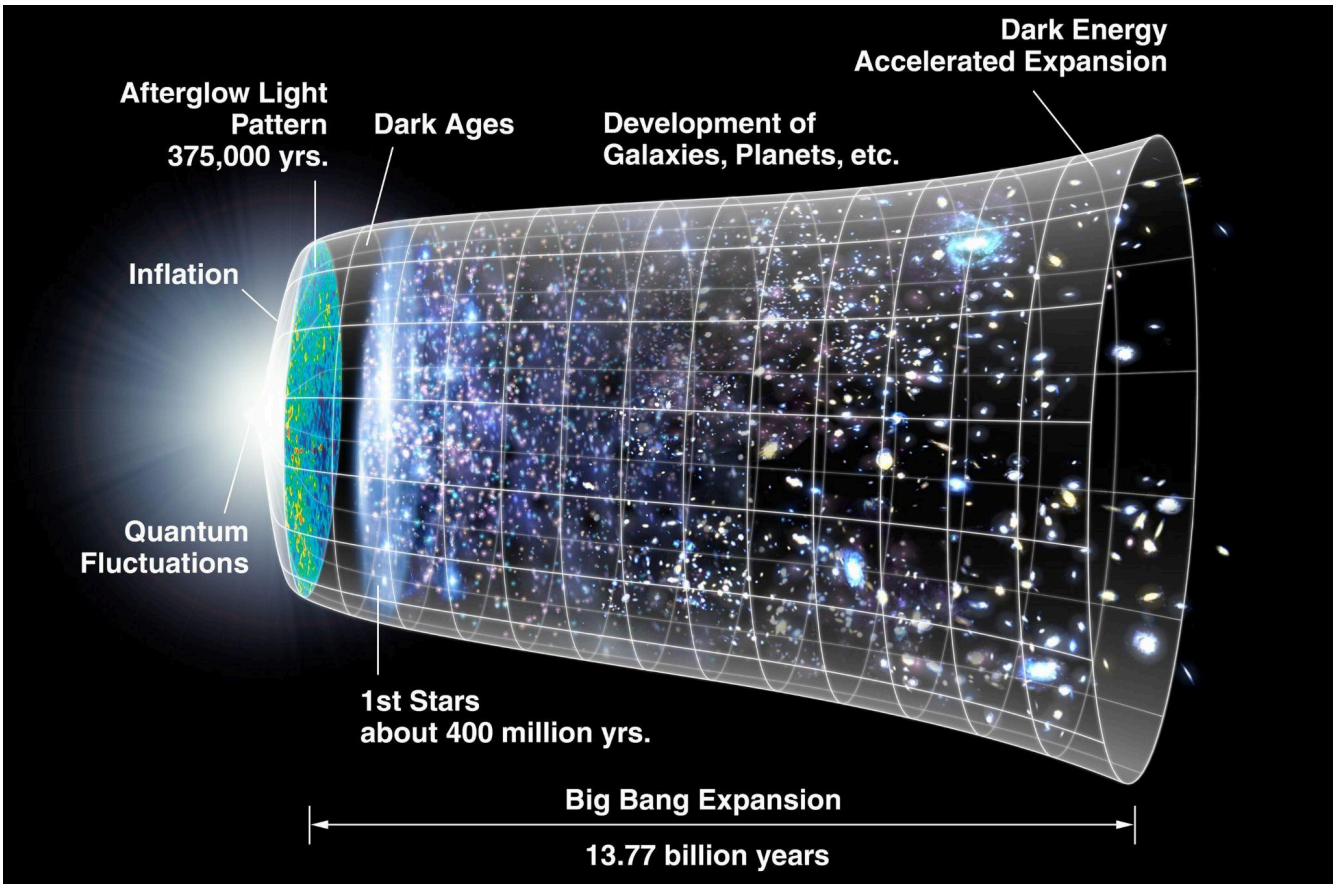
Learning Objective

Describe the various approaches you can take to brainstorm new theories and improve on existing theories during the coding process.

Reading the work of celebrated sociologists can be intimidating. How did they conceive of such provocative, nuanced, inspiring, and earth-shaking explanations for how the social world operates? Perhaps we are slightly overstating your personal admiration for sociological theory, but we think you will agree that arriving at compelling theories is a challenging task. The coding process we just described will move you far along this path if you invest sufficient time and effort into it. That said, there are a number of intellectual moves you can take to spark your creativity and guide you to potentially valuable insights.

We'll discuss three of these strategies: *neologisms*, *typologies*, and *concept mapping*. As you narrow your codes further and further through the open coding process we described earlier, you'll start to see the limits of existing theories. You'll encounter phenomena that are distinct from what past theorists have focused on, or are variations on those prior themes, but do not yet have a fitting name. You'll think of more precise or insightful ways of dissecting and categorizing social reality. And you'll come up with persuasive new models for the specific processes you've observed.

Neologisms



Neologisms don't necessarily have to be wholly original ideas you create in a sudden moment of genius. They can tap into the *folk* theories of your research participants, or they can recast existing theories with wording that is more intuitive and illuminating for others (including lay audiences). A classic example is the "big bang" theory of the origin of the universe. The term was actually coined by a critic of that theory, British astronomer Fred Hoyle. At the time, Hoyle recalled, he was arguing on news programs on behalf of an alternative explanation, the steady-state model: "I was constantly striving over the radio—where I had no visual aids, nothing except the spoken word—for visual images. And that seemed to be one way of distinguishing between the steady-state and the explosive big bang. And so that was the language I used" (Kragh 2013:2.29). *NASA/WMAP Science Team, via Wikimedia Commons*

One approach to generating new theory is by creating **neologisms**—new words or expressions, which in a research context would entail finding a precise, illuminating (and hopefully catchy) label for a particular phenomenon. These neologisms are often italicized in books and articles to emphasize that they represent a theorist's proposed way of understanding an aspect of reality.

Students are often a bit hesitant to put forward their own neologisms. They tend to want to defer to existing theory, stamping everything they see with their favored theorist's kosher labels. This isn't how science advances, however. As a critically minded and independent-thinking sociologist, you will want to question whether the prevailing theory—the prevailing term—actually encapsulates the complexity of social life in an accurate and thoughtful way. Don't be afraid to weigh into these debates, even if you are a novice researcher.

Now, it is very true that neologisms can be overdone. As you read through the literature on any topic, you'll inevitably

find yourself spotting a writer's italicized term and rolling your eyes. It may come across as awkward to you. Or you may feel existing terms cover the same ground more than adequately. When creating a neologism, you should keep these concerns in mind. For the first problem, the phrasing of the term, you might want to imagine yourself referencing the term in another writing context. Does it roll easily off the tongue? Does it need an explanation attached every time you use it, or can it potentially stand alone (the latter obviously being better)? Does it come across as gimmicky—a transparent attempt to exploit an issue briefly in the news, to be oh-so-cute in your self-important jargon, or to add hipness to the field (always a tragic endeavor)?

You should also ask yourself whether the literature really needs a new term for this phenomenon. Review any existing terms, comparing and contrasting them to your own and seeing if your new take adds something to the conversation. The answer will often depend on whether you've successfully uncovered a wrinkle in the existing theory—some aspect of the phenomenon that hasn't been sufficiently examined, and that therefore calls for a new label. One way to spot these shortcomings in current terminology is to diligently code the various dimensions of the topics or themes you observe in the ways we described earlier, which can flag specific areas the literature has neglected.

You should also note that you don't need to think up entirely new terms to advance the literature in your field. You can also contribute just by identifying elegant and instructive new ways to rework existing labels. A catchier neologism may replace the staid term currently being used. More important, if your new wording includes a vivid and accessible metaphor, it can help more people grasp the underlying idea. A classic (if somewhat contentious) example is the big bang theory, the now well-established view among physicists that the universe began in a state of unimaginably high density and temperature and then expanded outward. The Belgian priest and theoretical physicist Georges Lemaître proposed the idea of a universe spreading out from a single point, which he called the "primeval atom" (Soter and Tyson 2001). It was actually a critic of this theory, British astronomer Fred Hoyle, who coined the term "big bang." (Hoyle favored an alternative explanation for the universe's origins, the so-called steady-state model.) The "big bang theory" spread like wildfire within the scientific community and then the public at large—to the point of having a middling television show named after it. And you can see why: as Hoyle himself pointed out, the words "big bang" provided a ready visual image that made it easy for even lay audiences to get the gist of the theory. Nevertheless, scientists continue to gripe that the words "big bang" inaccurately imply that the universe's origins involved an explosion *in* space (rather than an expansion *of* space), and that sound somehow travels in space (Kragh 2013; Mathew 2013). For better or worse, the term has stuck.

Another alternative to thinking up whole new terms out of thin air is to borrow the terminology used by your research participants. As part of the aforementioned "folk theories" they cultivate, cultures and subcultures have unique ways of naming practices, beliefs, and other phenomena relevant to their members. You can advance existing theories just by bringing to light these examples of insider knowledge. Through the process of open coding, such terms will naturally stand out, and you should keep an eye out for how they might help explain what goes on in the settings you observe.



When you're considering a neologism, ask yourself if the new term improves on what is currently used in the literature—by capturing some aspects of the phenomenon of interest more fully, or by helping people understand a complex idea in a comfortable, even catchy way. If not, you should heed the sage advice of Regina George from *Mean Girls*: stop trying to make "fetch" happen. It's not going to happen. *Paramount, via Fandom*

One danger of this approach, however, is that you may naively adopt the perspective of your participants, who—as with any group—have particular moral and political outlooks that can blind them to other ways of looking at the world. They may not actually do what they say (or what their terms imply), and they may buy into dominant ideologies that could bias the way you interpret

their actions. An example of both the strengths and potential pitfalls of adopting folk terms from your participants is the "street" and "decent" terminology that ethnographer Elijah Anderson put forward in his book *Code of the Street* (2000). In the Philadelphia neighborhoods that Anderson studied, "street" was regularly used as a label for people more or less involved in criminal activities, whereas "decent" described upstanding citizens who stayed off the gang corners. These terms provided a powerful way of understanding how neighborhood residents navigated their social world. At the same time, they also garnered criticism for uncritically accepting moralistic and simplistic folk categories that "decent" folks appeared to be imposing on an unseemly "street" element—a charge that Anderson vehemently denied (Anderson 2002; Wacquant 2002).

Typologies



One or more interactive elements has been excluded from this version of the text. You can view them online here: https://viva.pressbooks.pub/sociology-research-methods/?p=371#oembed-1

Video 11.1. A Typology of Parenting Approaches. In this video, Annette Lareau (2011) discusses the typology she created to describe the parenting strategies that she observed among middle-class and working-class households.

As you break down the various dimensions of a concept in the open coding process, you will find yourself creating **typologies**—classification systems with distinct categories that capture different varieties of a given phenomenon. You can contribute to the scientific understanding of your topic in important ways by either devising a new way of typologizing a phenomenon or revising an existing typology to account for something it missed. A good typology helps us make sense of how a phenomenon varies and what factors might contribute to that variation. We've already mentioned one simple typology—Elijah Anderson's "street" and "decent" dichotomy—which categorized neighborhood residents according to their relationship with illicit activity, and which was drawn from the folk language of the culture Anderson observed. According to Anderson, the two groups could be distinguished by whether they upheld middle-class values ("decent") or defied them ("street"), conformed to or deviated from social norms, had traditional or nontraditional family structures, and lived private versus public lives. You could imagine how these various moral and cultural dimensions of "street" and "decent" life may have originated from descriptive codes created during the open coding process, which his analysis could have knitted together as dimensions of a broader cultural orientation toward the neighborhood and its criminal activity.

Another classic example of a simple but provocative typology is Annette Lareau's categorization of parenting strategies in her ethnography *Unequal Childhoods* (2011). Based on the observations that she and her colleagues conducted in households with children, Lareau argued that middle-class parents engaged in more intensive parenting—what she called "concerted cultivation." Like a gardener working with a delicate flower, they invested large amounts of time and effort into personally nurturing their children's academic and social skills. For instance, these more affluent and educated parents would fill children's schedules with enrichment activities, regularly engage in intellectually challenging, adult-like conversations with them, and model for them how to interact assertively with the authority figures in their lives. In contrast, working-class and poor parents pursued a strategy that Lareau called the "accomplishment of natural growth." These parents did not love their children any less, Lareau pointed out, but they believed—based on their own upbringing and culture—that it was better to let children roam free, without a parent stepping in to instruct them or intervene on their behalf.

The two parenting styles are summarized in **Table 11.1**. Let's consider how a typology like the one in the table might emerge during the coding process. Perhaps you might start this analysis with broad descriptive codes like "language use" and "intervention in institutions." Breaking down the wide variation in these observed activities could lead you to develop arguments about how parents of different social classes behave in different ways within these areas. Moving in the opposite analytical direction, you could also merge these codes. Recognizing variations across specific dimensions of parenting could prompt you to see them all as part of a larger class divide of more active and passive engagement with children. (We're only guessing at the analytical process that Lareau undertook, but you can listen to her describe for herself the parenting typology she developed in **Video 11.1**.)

Typologies can describe dimensions of any particular phenomenon—a range of diverse explanations, consequences, processes, and so on. It helps when typologies draw from an evocative and accessible metaphor, like the gardening imagery that Lareau used in her terms. You should also note that the categories in a typology typically represent **ideal types**—varieties of a phenomenon with particular characteristics imagined in their purest form. Using ideal types keeps a theory simple and streamlined. The real-world cases you observe may not fit perfectly within a single bucket, and some may span multiple ones. This is fine. In fact, it can inspire you to think of the reasons that certain cases straddle categories.

Another thing to keep in mind when constructing typologies is whether the concepts you're describing are being treated as independent variables or dependent variables. Are you trying to explain why the variation across dimensions is occurring, or do you want to understand the consequences of that variation? Oftentimes, the categories of your typology can be plausibly viewed as both causes and consequences, but for the sake of generating new theories, it can be helpful to be explicit about the orientation here you're taking.

Table 11.1. Typology of Parenting Approaches in Lareau (2011)

	Concerted Cultivation	**Accomplishment of Natural Growth**
Organization of Daily Life	Parents schedule numerous activities to occupy their children's time outside of school.	Parents provide children with large amounts of unstructured time.
Language Use	Parents negotiate with children at length and tolerate challenges to their decisions.	Parents do not allow for questioning of their directives.
Intervention in Institutions	Parents get involved to improve outcomes overseen by these institutions and also teach their children how to navigate such spaces.	Parents defer to authority figures and endure the consequences of disconnects between their views and those of the institution.
Overall Approach and Consequences	Parents are actively engaged in the enrichment of their children, who develop a sense of entitlement about the treatment they deserve.	Children develop their interests and talents relatively independently, often encountering constraints as a result the lack of adult support.

Note: Adapted from "Typology of Differences in Child Rearing" in Lareau (2011, table 1).

Obviously, you don't have to limit yourself to the sorts of binary or dichotomous typologies we've been discussing so far. You can easily identify three or more categories of a given phenomenon during your coding process (but note that adding more and more categories can clutter your theory and make it less easy to apply to other contexts). Furthermore, you can use **two-by-two tables** of the sort depicted in **Figure 11.9**. In this analytical approach, you tease out how

variation across two dimensions leads to distinct outcomes. This can help you understand how the categories you identified for your typology relate to one another.

In this example, Victor and his collaborators mapped out a typology of entrepreneurship based on the degree of creative independence (the extent to which entrepreneurs could conduct their work activities as they pleased) and the degree of financial independence (the extent to which they were obligated to financial backers) that people who started their own businesses enjoyed (Doody, Chen, and Goldstein 2016). More or less autonomy in these areas could situate an entrepreneur in one of the four categories described in the two-by-two table: a Silicon Valley entrepreneur (e.g., a tech startup founder reliant on venture capital funding), a Main Street entrepreneur (e.g., the owner of a mom-and-pop business that sustains itself on its own income), a corporate entrepreneur (e.g., an individual in a large company incentivized to create and manage new ventures), and a self-employed entrepreneur (e.g., an Uber driver or other gig worker constrained by the rules and oversight of a platform).

With a two-by-two table, you are explicitly contrasting the groupings you observe in your data according to how their characteristics in the two highlighted areas compare. You can imagine how such a comparative approach could naturally evolve from the process of coding similarities and differences across your cases and explaining why cases fall into particular groups (possibly ones that your participants have already identified as being important in their folk theories). Again, you can go beyond the simple categorization scheme in this example—having three or more rows or columns in your table, or locating groups somewhere on a continuum rather than discrete boxes—but this adds complexity that may make your theory unwieldy.

Economic relationships between market actors	**Creative independence (self-determination of product/service)**	**Creative dependence (employer or client determination of product/service)**
Financial dependence (focused on investor's expectations)	Silicon Valley entrepreneurship	Corporate entrepreneurship
Financial independence (focused on meeting personal income needs)	Main Street entrepreneurship	Entrepreneurial self-employment

Figure 11.9. Two-by-Two Table. In this two-by-two table from an article in *Sociology Compass* (Doody et al. 2016, figure 1), the authors present a typology of entrepreneurship. Two-by-two tables can illustrate in an accessible way how variation across two measures (here, financial independence across the rows, and creative independence across the columns) map onto different categories of a particular phenomenon (here, types of entrepreneurs). *Wiley*

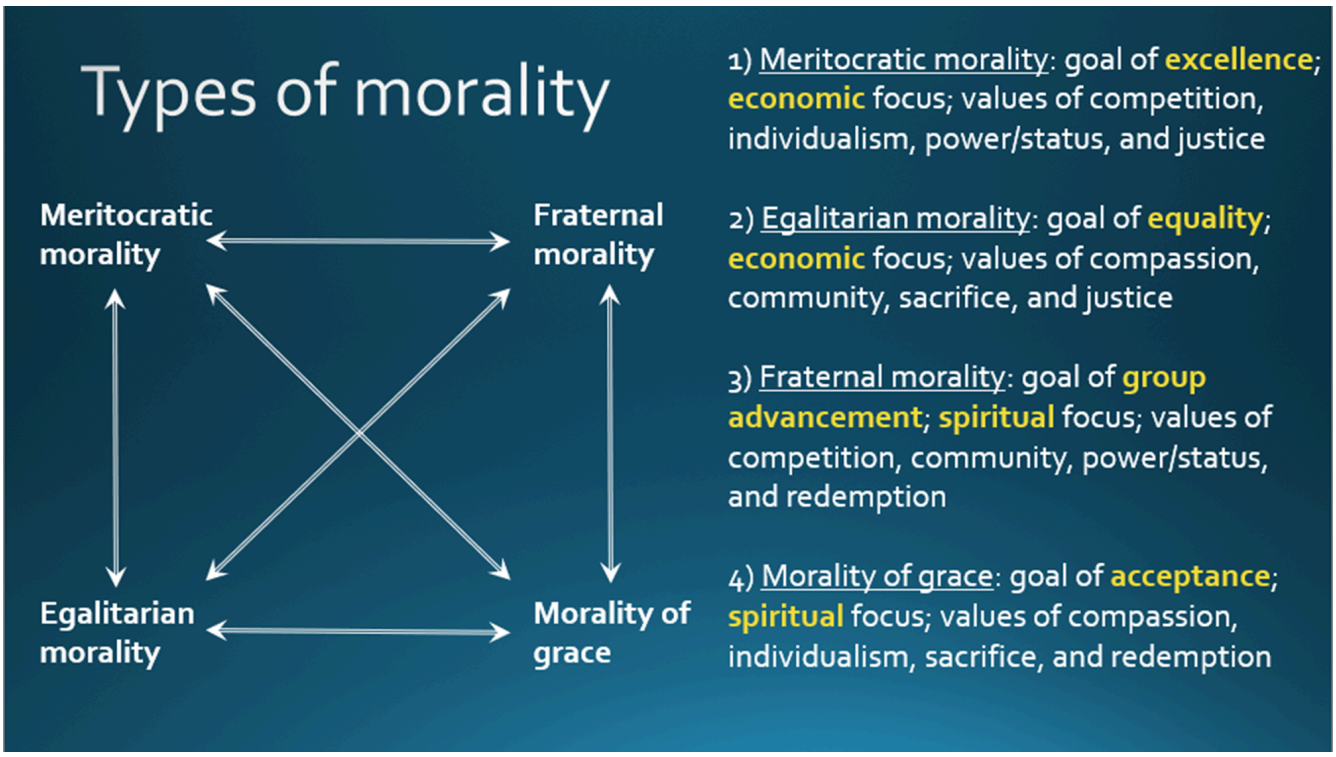


Figure 11.10. Complex Typologies. Beyond simple typologies and two-by-two tables, you can think of more complex categorization schemes that get at the relationships between different concepts or different dimensions of a single concept. For example, this typology from Victor's work compares and contrasts different moral perspectives (Chen 2015, 2021). Note that greater complexity, however, can overly complicate theories, which ideally are sufficiently lean and simple (parsimonious, to use the technical term) that they can be applied widely and easily.

If you want to get really convoluted, you can propose typologies of the sort described in **Figure 11.10**, which includes four moral perspectives and compares them across several dimensions. One thing to note about this approach is the importance of a clear visual representation of your theory. This goes for distinguishing between categories or dimensions within a typology as well as for presenting theorized relationships between concepts, which we will discuss next.

Concept Maps

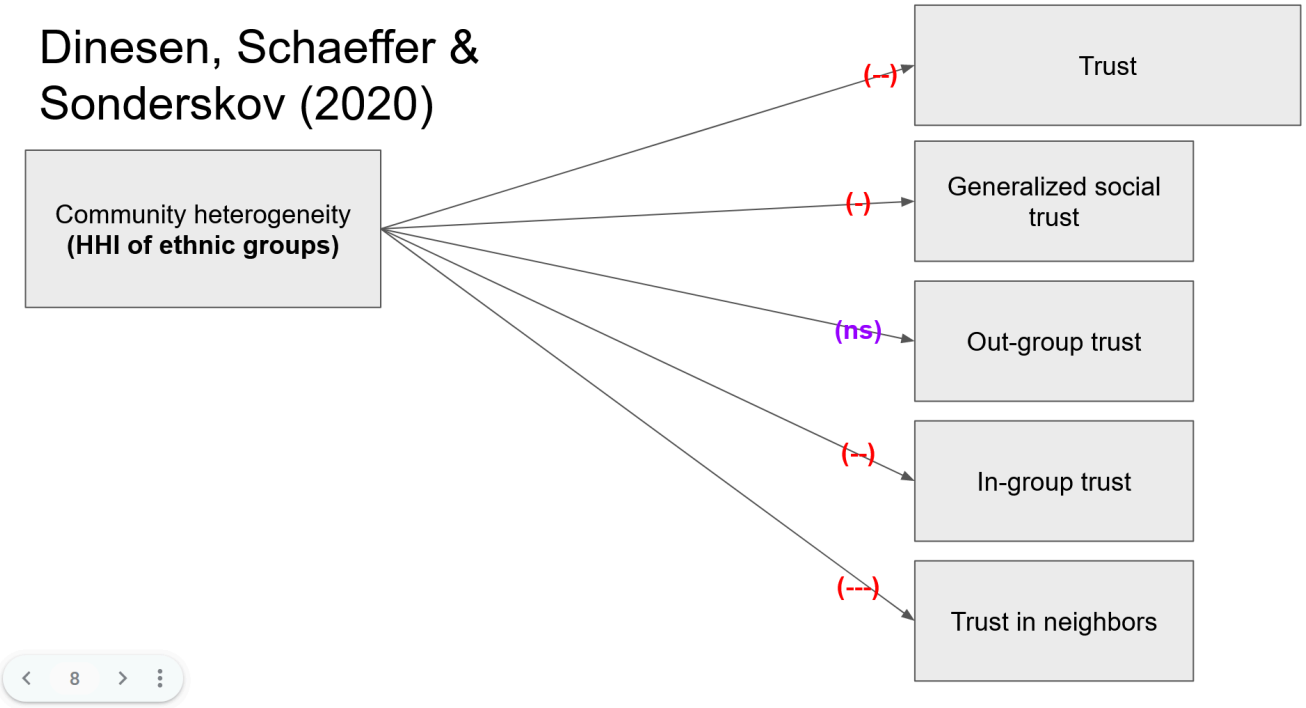


Figure 11.11. Concept Maps. Created for a literature review that Victor conducted, this concept map details the relationships being evaluated in a paper by the political scientists Peter Thisted Dinesen, Merlin Schaeffer, and Kim Mannemar Sønderskov (2020). Their meta-analysis incorporated the findings of 87 studies to determine whether the ethnic diversity of neighborhoods erodes reported levels of trust between individuals. This concept map summarizes the patterns of findings across the analyzed studies. Plus and minus signs refer to positive and negative relationships, and "ns" means that the relationship was not *statistically significant* (a concept discussed in Chapter 14: Quantitative Data Analysis).

In previous chapters, we've described how important concept mapping is to understanding existing theories and the relationships you plan to explore with your empirical research. Now we want to talk about how concept mapping can help you create or refine theories at the data analysis stage.

Concept mapping involves visualizing concepts and the relationships between them, generally using boxes and arrows (with the direction of the arrows showing the presumed direction of causality). You can lay out these concepts and relationships in a word processor like Microsoft Word or Google Docs or in a presentation program like Microsoft PowerPoint or Google Slides, which have simple-to-use tools for drawing and linking boxes and arrows. (You can also just start off with a piece of paper or a whiteboard.)

There are several ways to profitably use concept maps to create or improve upon theories. In **storylining**, for instance, you take the categories and relationships depicted in your concept map and use them to tell a **causal story** about the observed phenomenon. If sufficiently plausible, this story can be the central to your study's findings. (As we've discussed in previous chapters, you should do your due diligence in elaborating on any alternative explanations and not overstating the degree of certainty in this theory that your data justifies.)

You might also find it useful to compare the concept maps that emerge from your coding with any concept maps you generated for previous studies (see the example in **Figure 11.11**). What does your study include in its concept map that other studies do not? This difference becomes the basis for theoretical innovation. (You might even want to embark on this process of comparative concept mapping during the early literature review stage of your research, which we discussed in **Chapter 5: Research Design**.) If your concept map is exactly the same as that for another study, your study may still be contributing in a substantial way because of the specific methods it uses or the specific sample it has gathered. Visualizing your research in this way can help you pinpoint its contribution.

Key Takeaways

1. Neologisms may come to mind as you pour through your data and identify theoretical curiosities or empirical patterns in your participants' behavior and thought. They can advance the literature by providing a new way of defining and labeling an aspect of social reality or a more precise and intuitive way of characterizing existing theories.
2. Typologies help us understand the different dimensions of a given phenomenon. They can arise organically from the process of breaking down a concept or connecting different concepts during the coding process.
3. As you piece together a causal process through your coding, creating a concept map can help you see how the pieces fit together and understand what new features of the process your own work is uncovering or illuminating.

11.3. Writing Up Qualitative Results



A straightforward approach for writing up qualitative findings is to use the *sandwich method*, which moves back and forth between theoretical points (the buns) and empirical data (the meat). *Vilnis Husko, via Pexels*

Learning Objectives

1. Describe the sandwich method of writing up qualitative data.
2. Discuss the amount of descriptive detail you can and should include about your research participants to complement their quoted statements.

Now that you've analyzed your qualitative data, how do you write up your results? The bulk of your findings section will usually consist of quotations from your interviews or extended descriptions (excerpted in a more or less edited form)

from your field notes. It should be organized by theme rather than just being a random list of quotes or observations. In fact, it is often a good idea to break down the results section into subsections, with each theme explicitly noted in its subsection title. Within each subsection, you can then describe overall patterns in the data related to that theme as well as specific cases that back up your interpretations of what you observed or what your interviewees said.

There are a wide variety of ways to write up the paragraphs in your paper that describe patterns and cases, but we'll go over a staple approach, the so-called **sandwich method**.[1] Let's say you want to use an interview quote to make a particular point in your paper. The quote—the qualitative data—will be the meat (or vegan meat substitute) for your sandwich, surrounded by two buns of analytical context. If you are analyzing ethnographic observational data, then you would swap out that interview-based patty with an observation-based one (such as quoting the relevant field note or summarizing your observations in one or more sentences).

The "sandwich" is typically the description of a single case—say, a single interviewee or a single observed episode. Let's break down its layers:

1. The top bun starts at the **theoretical level**. To begin the paragraph in your write-up, you make a theoretical point, such as indicating that a particular pattern exists in your data.
2. Next, you move to the **empirical level**. The meat is the interview quotation or observational data you wish to include. At this level, you introduce the example of a particular participant, episode, or quote to back up your theoretical point (or complicate it). You will usually want to provide a description of any participant you mention or quote. This might include their name or pseudonym, age, gender, race, and any other relevant background details so that your reader knows something about the social position of the person you're referring to (more about this later). Either before or after that description goes the actual observational data, blockquote, or in-text quote or paraphrase.
3. The bottom bun returns to the **theoretical level**. At this level, you might want to reiterate your earlier theoretical point in different words, saying concretely how the example you just gave supports that point. You might provide some useful context for situating your example among the many episodes you observed or people you talked to. And you might discuss any exceptions or qualifications to your main point (e.g., explaining whether and why certain other cases deviated from the overall pattern). At some point, you will transition to the next case—perhaps to another "sandwich" entirely, or perhaps directly to the "meat" of a second example directly related to the first (e.g., an example that contradicts or complicates the point you just made). You can discuss the contrasts and comparisons you are making between cases, or you can pivot to an entirely new topic, perhaps starting a new paragraph to do so.

It's worth reiterating that the sandwich method is just one (formulaic) way of presenting the results of qualitative analysis, and many approaches are perfectly fine. For instance, sometimes ethnographers like to set the scene before making any theoretical points. They will have a paragraph or multiple paragraphs describing an episode or interaction—in other words, going right to the meat and skipping the top bun (the open-faced sandwich method, perhaps?). Then they will discuss that block of qualitative description in theoretical terms, indicating what broader point or points they believe it makes.

1. Victor thanks Katherine K. Chen for alerting him to this method and its name.



One useful strategy for writing up qualitative data is a case-*to-case comparison*: contrast two cases that are similar in many regards but different in one important area. Juxtaposing cases in this way can be good for explaining differences in outcomes based on that highlighted factor. When making a case-to-case comparison, you can ask yourself: How is the independent variable of interest decisive across your two cases? How do the causal mechanisms operating within these cases differ? *cottonbro studio, via Pexels*

One particularly good strategy that fits well with the sandwich method is to juxtapose two cases—perhaps two respondents or two settings—that are similar in many regards, but different in one key aspect. You would use the sandwich method to describe the first case, and then pivot to the second. This *case-to-case comparison* is good for theorizing how any differences in outcomes between the two cases may be related to the specific factor you singled out. You can make a case for how one independent variable, for instance, may be decisive across these cases, or how the causal mechanisms at work differ across these two cases. You can even structure your write-up like a double-decker sandwich: the single bun between the meat of the two cases can flag the key difference between them, and the bottom bun can focus on explaining the significance of that difference.

As we mentioned, the middle portion of the sandwich—your empirical data—can be in the form of a **blockquote**, an extended quotation that is indented (typically by a half-inch) to highlight it. In addition to interview quotes, blockquotes can feature excerpts from field notes, though it is usually more straightforward to include that observational material directly in the text rather than setting it off in this way. When you indent quoted material as a blockquote, you should not put quotation marks around it. The blockquote's text should be single-spaced even if the rest of the text is double-spaced, and the text size can be set slightly smaller. Any parenthetical citation or attribution goes *after* the period or other punctuation that ends the blockquote, as shown in **Figure 11.12**, and it does not require its own concluding period.

Seth, 28, a "somewhat liberal" unemployed health care office assistant, echoed this perspective, explaining that media accounts were especially dismal so that the outlets could "make money," explaining,

> fear drives viewership, and especially with people being at home with nothing to do except watch the news. All the news is about, it's like everything else stopped happening, almost. It's virus, virus, virus. I think they found something that they could use to drive their own viewership and I think they're taking advantage of it.

Contributing to claims of sensationalism were complaints that the news was intentionally intended to depress the audience and that anything positive was ignored or hidden. While the "good news" was when someone survived the virus, participants complained that there was very little focus on COVID-19 survival statistics:

> From all of the bad stuff I heard, I heard it twice saying that the majority of people can [survive] COVID. . . . You get it, but then most people have beat it. Why aren't you not saying, why are you not balancing it out? You're just talking about all of the negative. (Tyler, 33, former personal trainer and Democrat)

> I think that they're milking it. They're milking it, and it's kind of like they want to depress you. They're trying to make everything worse, at least from what I see. I know it's bad, but I'm saying I feel like they're making everything worse than it really is just for bigger news and everything else, to make you even more nervous than you already are. It seems like all the news channels hate Trump. So they seem to make it worse because then he'll get blamed for it. But it's affecting us because you're depressing us. (Carlos, 43, Democrat and former restaurant server)

Figure 11.12. Blockquotes in Qualitative Studies. In their study of Americans' distrust of the media during the Covid pandemic, Alexandrea J. Ravenelle, Abigail Newell, and Ken Cai Kowalski (2021:5) quote their respondents at length using blockquotes. Note the different ways that the quotations above are attributed, with the top blockquote introducing the respondent before the blockquote and the bottom quotes including such details in parentheses at the end of each quote. *Sage*

Writing styles have different conventions about how long a piece of text needs to be before you should use a blockquote. The style guidelines for the American Psychological Association (APA) and American Sociological Association instruct authors to place quotations in a block when they are more than 40 words, whereas Chicago style has a 100-word threshold. However, you should use your judgment about when a blockquote is called for. A downside of blockquotes is that your theoretical point may get muddled when the excerpted text is too long. Too many blockquotes can also make your writing seem dull and formulaic. It may be better to use a mix of direct and indirect quotes and paraphrases to convey the same information.

Here are some general pointers about how to write up qualitative data that apply to both interview and observational data:

1. **Use past or present tense consistently.** Using either tense is fine when describing a scene or attributing a quote, but unless it is really necessary, try not to switch back and forth between tenses, which can be disorienting for the reader.
2. **Avoid using distracting attribution words in quotes.** For the most part, you can use "said" (or "says") and "replied" (or "replies") exclusively. These words don't draw as much attention to themselves, which means your readers will

focus on the points you're making rather than your flashy prose.

3. **Put the attribution after the quote.** Unless you're using a blockquote (discussed later), it generally sounds better to put the attribution after the quote ("*I love sociology*," *she said*) rather than before it (*She said*, "*I love sociology*"). If you have a quote that is more than two sentences long, you can put the attribution between those two sentences ("*I love sociology*," *she said. "It's so much sexier than economics.*"). Of course, mixing up the format for attributions also works fine and can make your writing livelier. If you're writing for a publication that uses American (as opposed to British) English, then make sure to put any punctuation within the quotation marks rather than outside it—with the exception of quotation marks and exclamation points that are actually part of the quote.
4. **Refer to "interviewees," "respondents," or "participants."** As we mentioned in **Chapter 10: In-Depth Interviews**, the term "interviewees" probably sounds the least awkward to lay readers, so you probably want to stick with that if you're interested in making your work more accessible. With that goal in mind, we'd also suggest avoiding more stilted terms like "interlocutors" or "subjects." The catchall term "research participant" or just "participant" is fine to use when talking about someone you observed but did not interview.

One useful practice that ethnographers have adopted for writing up results is to reserve quotation marks only for interview data whose wording they are certain about—that was recorded and transcribed, for instance. When recounting quotations from memory or quickly jotted-down notes (sometimes the only options available when observing in the field), researchers will use italicized text to note that the quotation may not be a precise word-for-word transcription of what the person said.

Writing Up Interview Results



Observing families in their homes can generate incredibly rich and evocative descriptions about their lives, allowing ethnographers to draw not just on what participants tell them but also on the suggestive details of their appearance and mannerisms—as Annette Lareau did to great acclaim in her ethnography *Unequal Childhoods* (2011). Yet this immersive kind of ethnography can also cause issues, such as the backlash that Lareau received from one of the families she profiled. Among other things, the Handlon family was upset by descriptions of their house as messy and their daughter as "chubby." *August de Richelieu, via Pexels*

When you are writing up material from in-depth interviews, there are several additional issues to consider. First, you will have to find a way to introduce each interviewee and provide a few details about their background to put what they say in context. The following information typically goes into that brief introductory description:

1. **Pseudonym.** If you have promised confidentiality to your participants, which is typically the case, then you obviously can't use their real names. That said, we'd suggest using a made-up name rather than just "Respondent A" or "Interviewee 17." A made-up name is much more memorable for the reader than a generic label, and pseudonyms are especially useful for interviewees who are more important to your research and you plan on quoting multiple times. On the other hand, if you are quoting a respondent only once, it's also acceptable to leave out a label altogether (e.g., by simply describing the person as a "a 20-year-old African American female student" and including their quote).
2. **Background details.** Generally speaking, it is a good idea to include some background about an interviewee when

you introduce them in your paper, so that the reader knows something about their social position in relation to other respondents (and therefore how to put their words in context). Age, gender, and race are fairly standard to include, given their importance as social markers in our society—for example, "Samantha, a 32-year-old Latina woman from Richmond." Other details like the respondent's occupation or position within an organization, their socioeconomic status (education level, self-identified social class, or even income level), and their city of residence are also common. That said, you should include whatever descriptors are most relevant to your study. For instance, if you are doing a study that examines the attitudes of various sexual minorities, you might want to mention the sexual orientation of each respondent. Note that you do not have to include all of these details in a single sentence, which can make your writing clunky. As shown in **Figure 11.13**, you can instead spread out relevant details across the sentences of a paragraph. For blockquotes, you may want to put the identifiers in parentheses at the end of the quotation, especially if you are including quotes from different respondents whom you need to distinguish. For instance, "(Respondent A, 23 years old, working-class African American male student)" would appear at the very end of that individual's quote, even if the blockquote was multiple paragraphs long.

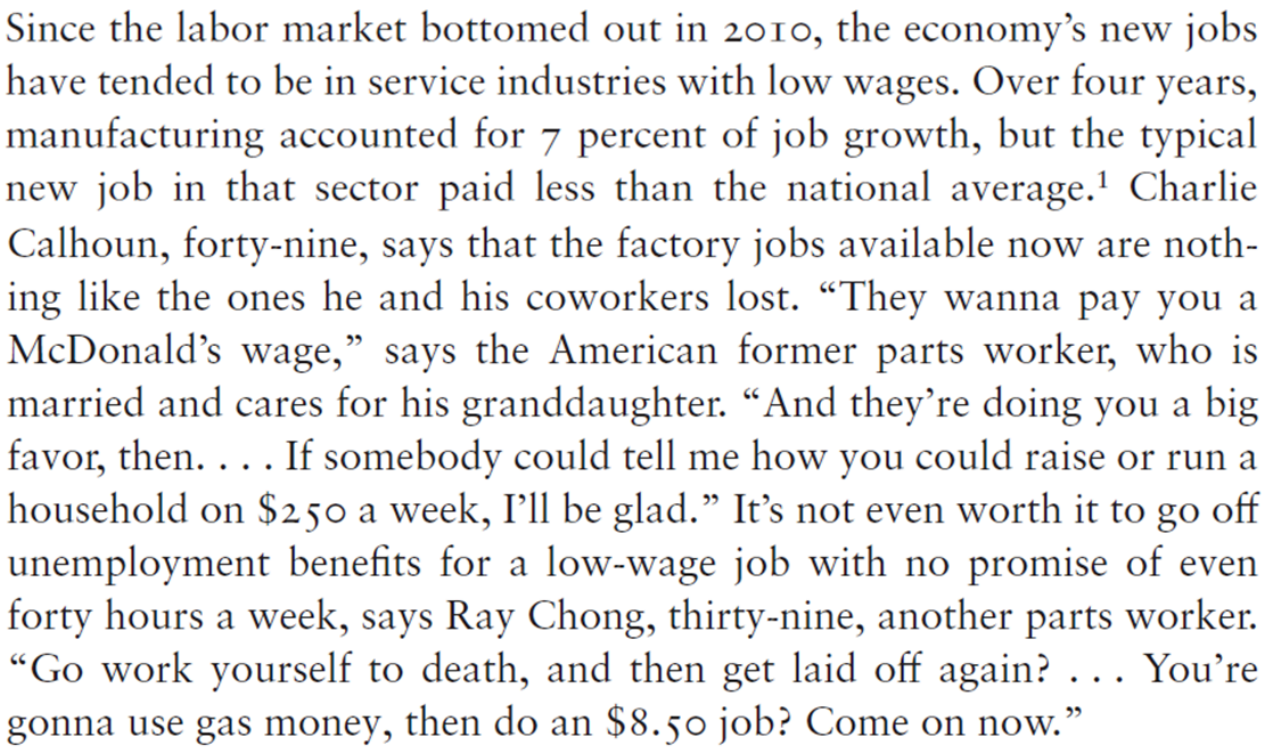

Since the labor market bottomed out in 2010, the economy's new jobs have tended to be in service industries with low wages. Over four years, manufacturing accounted for 7 percent of job growth, but the typical new job in that sector paid less than the national average.[1] Charlie Calhoun, forty-nine, says that the factory jobs available now are nothing like the ones he and his coworkers lost. "They wanna pay you a McDonald's wage," says the American former parts worker, who is married and cares for his granddaughter. "And they're doing you a big favor, then. . . . If somebody could tell me how you could raise or run a household on $250 a week, I'll be glad." It's not even worth it to go off unemployment benefits for a low-wage job with no promise of even forty hours a week, says Ray Chong, thirty-nine, another parts worker. "Go work yourself to death, and then get laid off again? . . . You're gonna use gas money, then do an $8.50 job? Come on now."

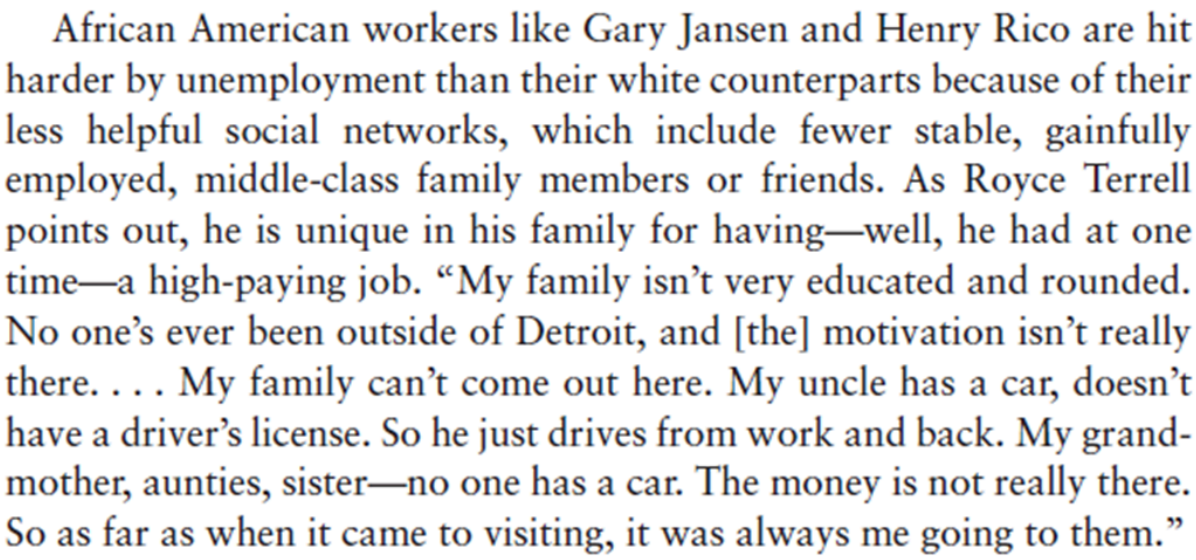
Figure 11.13. Introducing Interviewees. When mentioning research participants (especially interview respondents) in a paper, you generally should include basic demographic details to give readers a sense of their social positions and therefore how best to contextualize their comments or actions. Note, however, that you don't have to dump all that information in one long label. As in the example excerpt (Chen 2015:43–44), you can spread out those details throughout the sentences in a paragraph. *University of California Press*

How should you write up your interviewee's actual words? Again, there is no "right" way, but we would advise that you use full quotations—whether blockquotes or not—only when you have an especially expressive, eloquent, or theoretically compelling quote that you want to highlight, or when it is important that you present a longer quote to provide context. Partial quoting (of a few words or phrases) or paraphrasing is appropriate for conveying more straightforward

information, and this approach will ensure that your voice and your theoretical arguments are not lost amid your respondents' words. Save up the direct quotes for powerful and provocative material.

On a related note, including lots of interviewees in your write-up can clutter your analysis, especially if they are saying more or less the same things. While it can be important to convince readers that you have not cherry-picked your respondents and actually did talk to a wide range of people, you don't want your paper to become a wall of quotes from faceless people. Prioritizing key cases and vivid quotes will play to the strengths of qualitative research and allow your theoretical points to come across clearly and indelibly. In fact, some qualitative researchers will choose in their articles or books to showcase a very small group of exemplary cases—say, three to six interviewees. The analysis will return again and again to the stories of these *focal participants* or *key informants*, as if they were the main characters in a play. Like bit players, other respondents will appear only occasionally, their data occasionally sprinkled in to complement a particular point being made. This focused approach to writing up qualitative data (used almost exclusively in trade books meant for popular audiences) can help ground the narrative and make it much more accessible to readers.

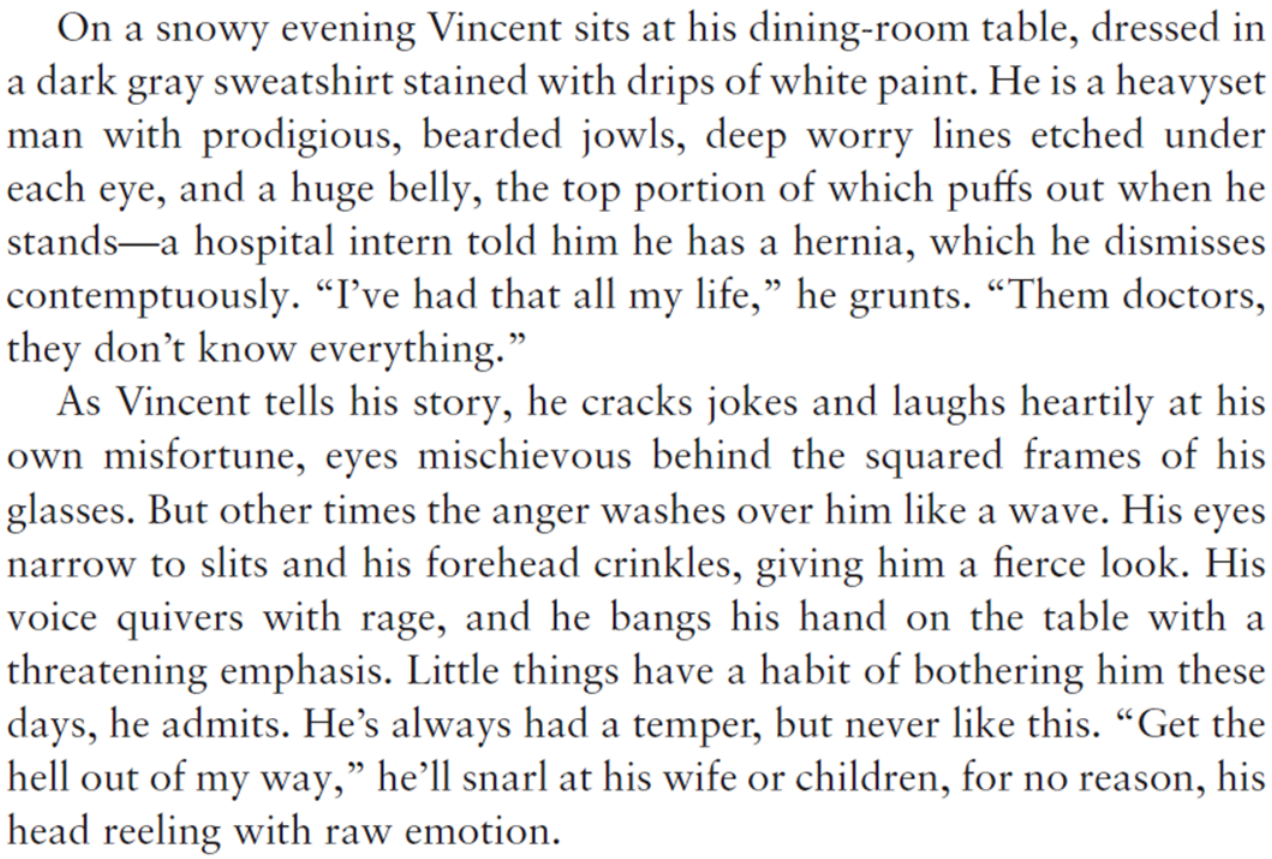
When quoting people, it is generally acceptable to make slight edits to their wording in order to more clearly convey what they meant to say. So long as you are not altering their intended meaning, for instance, you can fix minor grammatical errors or add any words that they inadvertently left out. In these cases, you normally want to put the inserted text in brackets and indicate any cut text with an ellipsis. You can see an example in **Figure 11.14**. Note that writers sometimes use an ellipsis enclosed in brackets to signal that text was removed (as opposed to the respondent pausing in their speaking). This added step is not strictly necessary, however, and it will make your writing seem more technical. You also don't need to use brackets or ellipses when you remove so-called filler words, meaningless phrases that people naturally insert into their speaking as they think through a response: "um," "like," "you know," "ah," "hmm," "right," and "alright," for instance. That said, you may want to leave a few filler words in your quotes, which can make the language livelier and also offer subtle clues about what the person was thinking or feeling when they said what they said.



African American workers like Gary Jansen and Henry Rico are hit harder by unemployment than their white counterparts because of their less helpful social networks, which include fewer stable, gainfully employed, middle-class family members or friends. As Royce Terrell points out, he is unique in his family for having—well, he had at one time—a high-paying job. "My family isn't very educated and rounded. No one's ever been outside of Detroit, and [the] motivation isn't really there. . . . My family can't come out here. My uncle has a car, doesn't have a driver's license. So he just drives from work and back. My grandmother, aunties, sister—no one has a car. The money is not really there. So as far as when it came to visiting, it was always me going to them."

Figure 11.14. Editing Interview Quotes for Clarity. Unless you are engaged in the kind of detailed conversational analysis where every misspoken word or verbal tic is important to capture, you can usually do slight edits of the quotations you use to fix minor grammatical errors and remove filler words like "um" and "like." You can note such changes with brackets and ellipses, respectively, as in the example excerpt (Chen 2015:108). *University of California Press*

We'll wrap up this discussion of presenting interview data with some additional thoughts on providing background details about your respondents. Victor—who, as previously noted, was originally trained as a journalist—likes to provide additional information about each interviewee, including describing their physical appearance, to flesh out their individual portrait in a published text. You can see what this looks like in **Figure 11.15**. With your institutional review board's stipulations in mind, you will need to decide on what you find appropriate in terms of the level of detail and context you provide. You obviously need to honor any promises of confidentiality you gave to your respondents, and including too much information may risk violating that confidentiality. Furthermore, providing details like physical appearance can come across as unflattering or even creepy. Ethnographers sometimes regret even an offhand, seemingly innocuous comment that, they later learn, a research participant felt slighted by (Lareau 2011).



On a snowy evening Vincent sits at his dining-room table, dressed in a dark gray sweatshirt stained with drips of white paint. He is a heavyset man with prodigious, bearded jowls, deep worry lines etched under each eye, and a huge belly, the top portion of which puffs out when he stands—a hospital intern told him he has a hernia, which he dismisses contemptuously. "I've had that all my life," he grunts. "Them doctors, they don't know everything."

As Vincent tells his story, he cracks jokes and laughs heartily at his own misfortune, eyes mischievous behind the squared frames of his glasses. But other times the anger washes over him like a wave. His eyes narrow to slits and his forehead crinkles, giving him a fierce look. His voice quivers with rage, and he bangs his hand on the table with a threatening emphasis. Little things have a habit of bothering him these days, he admits. He's always had a temper, but never like this. "Get the hell out of my way," he'll snarl at his wife or children, for no reason, his head reeling with raw emotion.

Figure 11.15. Extended Descriptions of Interviewees. As illustrated in this book excerpt (Chen 2015:87), some sociologists include detailed descriptions of their respondents in their published work. The rationale is that capturing physical appearance and mannerisms can provide useful clues about individuals' social positions and contextualize their spoken comments—in addition to making for more engrossing reading. However, such an ample amount of detail is not necessary for most interview-based studies, and you will need to consider your interviewees' privacy and institutional review board guidelines when including information beyond their quotes. *University of California Press*

Many interview-based studies rely wholly on the words of their respondents as data, not using any observational information, and that is perfectly fine. If you were doing interviews for a program evaluation, you probably wouldn't even want to gather nonverbal detail: you just want their opinions about the program, and it would be awkward to include more. That said, as we discussed in **Chapter 10: In-Depth Interviews**, observational data can add depth to the spoken interview data, providing vital clues about a respondent's social position and their attitude or orientation toward the things they say. If we focus just on the words alone, we may miss valuable signals about how people hedge their

answers, convey passion about a particular topic, feel ambivalent or apathetic about the particular line they're giving the interviewer, or harbor internal contradictions and tensions that they may not express in their language (Pugh 2013). Likewise, visual descriptions of people's appearance and their surroundings can situate the reader in their social realities to the point that they do not come across merely as disembodied voices saying words. For instance, clothing can tell us something about their socioeconomic status, while expressions of exhaustion and disgust can convey the underlying moods that color their responses to particular questions. All this additional information helps us with **triangulation**, introducing complementary and potentially confirmatory data that allows us to vet and expand upon what we learn through people's spoken statements.

Ultimately, good qualitative work is storytelling, and one way you make your qualitative cases vivid, memorable, and theoretically rich is through detailed narrative and description. Sometimes respondents appreciate this immersive approach, which helps them feel fully heard and seen in the published text. And even if you don't wind up using all the details you gather, it can just be good practice to record in your field notes as much information as possible—verbal, visual, and otherwise—so that you know you are making full use of the qualitative method, which is unique in allowing for this flexibility.

Key Takeaways

1. The sandwich method takes a specific empirical case (or group of cases) and contextualizes it theoretically. It starts by presenting an overall pattern or argument, providing empirical support for that point, and then either reiterating or building on that point, often in reference to other cases.
2. When writing up interview data, you should provide context about your respondents by including demographic and other pertinent details whenever possible. If appropriate, you may also wish to complement their quotes with descriptions of their appearance, mannerisms, and actions, which can provide clues about their social position, their current state of mind, and other useful background information.

11.4. References

Anderson, Elijah. 2000. *Code of the Street: Decency, Violence, and the Moral Life of the Inner City*. Reprint ed. New York: Norton.

Anderson, Elijah. 2002. "The Ideologically Driven Critique." *American Journal of Sociology* 107(6):1533–50. doi: 10.1086/342772.

Becker, Howard S. 1998. *Tricks of the Trade: How to Think about Your Research While You're Doing It*. Chicago: University of Chicago Press.

Chen, Victor Tan. 2015. *Cut Loose: Jobless and Hopeless in an Unfair Economy*. Oakland: University of California Press.

Chen, Victor Tan. 2021. "The Mirage of Meritocracy and the Morality of Grace." Pp. 58–72 in *The Routledge Handbook on the American Dream*. Vol. 1, edited by R. C. Hauhart and M. Sardoč. New York: Routledge.

Chen, Victor Tan, and Jesse Goldstein. 2024. "The Theatre of Entrepreneurship: Learning to Perform the Speculative Self in University Entrepreneurship Programs." *Social Problems* 71(1):203–19. doi: 10.1093/socpro/spac008.

Corbin, Juliet, and Anselm Strauss. 2007. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. 3rd ed. Los Angeles: SAGE.

Dinesen, Peter Thisted, Merlin Schaeffer, and Kim Mannemar Sønderskov. 2020. "Ethnic Diversity and Social Trust: A Narrative and Meta-Analytical Review." *Annual Review of Political Science* 23(1):441–65. doi: 10.1146/annurev-polisci-052918-020708.

Doody, Sean, Victor Tan Chen, and Jesse Goldstein. 2016. "Varieties of Entrepreneurial Capitalism: The Culture of Entrepreneurship and Structural Inequalities of Work and Business Creation." *Sociology Compass* 10(10):858–76. doi: 10.1111/soc4.12407.

Esterberg, Kristin G. 2001. *Qualitative Methods in Social Research*. Boston: McGraw-Hill.

Illich, Ivan. 1968. "To Hell with Good Intentions." Presented at the Conference on InterAmerican Student Projects (CIASP), April 20, Cuernavaca, Mexico.

Kragh, Helge. 2013. "Big Bang: The Etymology of a Name." *Astronomy and Geophysics* 54(2):2.28–2.30. doi: 10.1093/astrogeo/att035.

Lareau, Annette. 2011. *Unequal Childhoods: Class, Race, and Family Life*. 2nd ed. Berkeley: University of California Press.

Lofland, John, David A. Snow, Leon Anderson, and Lyn H. Lofland. 2005. *Analyzing Social Settings: A Guide to Qualitative Observation and Analysis*. 4th ed. Belmont, CA: Wadsworth.

Martin, Leslie. 2008. "Boredom, Drugs, and Schools: Protecting Children in Gentrifying Communities." *City and Community* 7(4):331–46. doi: 10.1111/j.1540-6040.2008.00270.x.

Mathew, Santhosh. 2013. *Essays on the Frontiers of Modern Astrophysics and Cosmology*. New York: Springer.

Miles, Matthew B., A. Michael Huberman, and Johnny Saldaña. 2019. *Qualitative Data Analysis: A Methods Sourcebook*. 4th ed. Los Angeles: SAGE.

Mitchell, Tania D. 2008. "Traditional vs. Critical Service-Learning: Engaging the Literature to Differentiate Two Models." *Michigan Journal of Community Service Learning* 14:50–65.

Pugh, Allison J. 2013. "What Good Are Interviews for Thinking about Culture? Demystifying Interpretive Analysis." *American Journal of Cultural Sociology* 1(1):42–68. doi: 10.1057/ajcs.2012.4.

Ravenelle, Alexandrea J., Abigail Newell, and Ken Cai Kowalski. 2021. "'The Looming, Crazy Stalker Coronavirus': Fear Mongering, Fake News, and the Diffusion of Distrust." *Socius* 7:23780231211024776. doi: 10.1177/23780231211024776.

Soter, Steven, and Neil deGrasse Tyson, eds. 2001. *Cosmic Horizons: Astronomy at the Cutting Edge*. New York: New Press.

Wacquant, Loïc. 2002. "Scrutinizing the Street: Poverty, Morality, and the Pitfalls of Urban Ethnography." *American Journal of Sociology* 107(6):1468–532. doi: 10.1086/340461.

12. EXPERIMENTS



When you think of the word "experiment," what comes to mind? Maybe you associate the term with images of white-coated scientists pouring chemicals into test tubes. And yes, without a doubt, experiments are central to scientific inquiry of all kinds. But it's worth considering how various kinds of experiments pervade modern life. Marketing and social media companies run experiments all the time to figure out what products we are more likely to buy or what links we are more likely to click on. We watch television shows that use experimental methods to vet the effectiveness of different products or test whether urban legends are actually true. And we each "experiment," in a broad sense, as we go about our daily lives. We try out a new energy drink or hot sauce, in search of something better. We compare different workout regimes or meditation techniques and take note of how we feel afterward.

This chapter is aimed at giving you a richer sense of how scientists use experimental methods to learn about the social world. Fundamentally, an experiment involves the researcher introducing a variable (the **independent variable**, also known as the *experimental stimulus* or *treatment*) and then observing what happens to another variable afterward (the **dependent variable**, also known as the *outcome variable*, *outcome measure*, or *response variable*). When we engage in experiments, we are typically pursuing a deductive approach (as discussed in **Chapter 4: Research Questions**), asking, "What effect does the independent variable have on the dependent variable?" Compared with other research methods, experiments excel at answering the "effect" part of this question. They tell us convincingly whether a cause-and-effect relationship really exists between two variables.



The hosts of the show *MythBusters* used their wide-ranging technical skills to test widely believed claims and replicate legendary stunts—from whether it would be faster to weave between lanes of traffic or drive in just one, to whether a ninja could breathe underwater using a blow gun and then assassinate someone with a dart. While obviously not set up in a scientifically rigorous fashion, the show's "experiments" applied elements of the scientific method in an entertaining—and often explosive—fashion. *Discovery, via NPR*

In judging cause and effect, experiments are often considered to be the "gold standard" of scientific research design, the yardstick against which all others are judged. When they are conducted with all the safeguards we describe in this chapter, we say that experiments have high **internal validity** (not to be confused with the validity of a *measure*, which we discussed in **Chapter 7: Measuring the Social World**). Internal validity is our study's ability to determine if changes in the independent variable truly cause changes in the dependent variable. When we have high internal validity, we are able to rule out any **confounders**—variables other than stimulus or treatment that may be influencing the dependent variable.

In an experiment, we conduct research in a setting where we can manipulate the independent variable directly. Under properly controlled conditions, we can then see what happens to the dependent variable—all other things being equal—after the independent variable is changed. This approach produces results with higher validity than **observational data**, information gathered without the use of experimental procedures (i.e., where the researcher only "observes" and cannot alter the independent variable).

Most sociologists use ethnographic, survey, or interview data rather than experimental data, given the difficulties of studying social phenomena in a lab setting. Social psychologists are a major exception to this generalization: you can think of them as the perspicacious love child of two disciplines—employing the laboratory techniques foundational to psychology while studying the topics and issues foundational to sociology. Furthermore, as we will describe, sociologists

can use observational data in **quasi-experiments** that *approximate* a true experimental design, improving their study's internal validity beyond what they could do with merely an analysis of observational data.

Even if you do not wind up using experiments in your own research, it is essential that you understand why a well-done experiment excels at **causal inference**—its ability to infer, or conclude, whether a relationship is truly causal—and why other methods fall short. To fully appreciate the high internal validity of experimental approaches, let's review the four criteria for causal theories that we mentioned way back in **Chapter 3: The Role of Theory in Research**:

1. **Correlation:** A change in one variable is associated with a change in the other. (This is the bare minimum requirement for causality—remember, "correlation is not causation"!)
2. **Temporal precedence:** The cause occurs before the effect. In this case, we're interested in the *sequence* of events: the change in the independent variable has to take place before the change in the dependent variable, or the former cannot logically be considered a cause of the latter.
3. **Rejection of alternative hypotheses:** After testing whether other variables explain the phenomenon of interest, we conclude that these other possibilities are less convincing.
4. **Causal mechanisms:** We can describe the specific process by which changes in the independent variable lead to changes in the dependent variable.

Experiments can be designed to rigorously test whether an observed relationship meets the first three of these criteria. Their results can also shed light on the fourth criterion, *causal mechanisms*, which we typically describe by using theory—that is, applying logic and our knowledge of past research to understand how the variables of interest might be connected. (As we noted in earlier chapters, qualitative research excels at sketching out these causal processes.)

The first two criteria, *correlation* and *temporal precedence*, are a relatively low bar for a research study to meet. As we'll describe in **Chapter 14: Quantitative Data Analysis**, we can even use statistical analysis to generate a number (e.g., a **correlation coefficient**) telling us how strongly associated our two variables are. However, that's not enough for us to be able to conclude that a change in the independent variable causes a change in the dependent variable. We still need to show that the results we see are due solely to the influence of the independent variable and not to any other factor.

That's the hardest of the three standards to meet—what we call **nonspuriousness**. You might remember that **Chapter 3: The Role of Theory in Research** discussed how we can easily be fooled by spurious relationships if we don't account for all confounding variables, which are associated with both the independent and dependent variables of interest and therefore might be the real reason we detect a correlation. Experiments can help us rule out any confounders and avoid **spuriousness**. If we set up the experiment properly, we can eliminate all alternative explanations that might cause the outcomes we observe other than our independent variable. But that requires us to have all the necessary pieces of a rigorous experiment in place: pretests and posttests, experimental and control groups, random assignment to either group—that is, all the features of the ideal, or "true," experiment that we will describe throughout this chapter. Any missing piece will open the door to other possible **causal stories** that explain our experiment's results.

In addition to internal validity, we also often talk about **external validity** when we critique an experiment's design. If internal validity is about *causality*, external validity is about generalizability—specifically, whether the results of an experiment conducted on a particular sample can be generalized to other situations and populations beyond the context of our research setting. While experiments tend to have high internal validity, they generally do worse with external validity. First of all, a **laboratory experiment** is typically done in an office, with researchers trying their best to emulate conditions in the real world. This artificial setting may not adequately match what actually happens in everyday social life. To improve upon the contrived nature of the lab setting, scientists may use **field experiments**, experiments conducted in real-world settings. Another issue is that we typically don't have large random samples when we conduct experiments. Remember from our discussion in **Chapter 6: Sampling** that the standard method of obtaining a representative sample is to choose the cases in that sample randomly from the target population. The larger our

sample is, the more precise our results are. That approach of generating large samples (a large n) is not typically possible or desirable with experiments.

We'll have more to say about field experiments later, but we want to highlight here the second problem, which is harder to address. Randomly selecting participants for a social experiment is not always possible. Since we are ethical social scientists, we can't force people to be part of our study. We must rely on volunteers, who can be difficult to recruit in large numbers. External validity is therefore a perennial problem with experimental designs, and we'll go into some of the related issues in detail.

Opening chapter image credit: *Martin Lopez, via Pexels. Adapted by Bizhan Khodabandeh.*

12.1. Pre-experimental Designs: What You Should Not Do



Experimental designs that fail to include all the features of a true experiment—including pre- and posttests and randomly assigned control and experimental groups—usually suffer from problems with *internal validity*. We have much less certainty that a change in the independent variable truly causes a change in the dependent variable. *Clint Patterson, via Unsplash*

Learning Objectives

1. Discuss the various types of threats to internal validity when a control group is not used in an experiment.
2. Define selection bias and describe how the random assignment of members to control and experimental groups can address this problem.

Somewhat ironically, a good strategy for learning about experimental design is to start with designs that, for various reasons, just don't meet the requirements we discussed earlier for concluding that a causal relationship exists between two variables. Technically, we call these approaches **pre-experimental designs**, because they have not yet achieved a "true" experimental design that is internally valid. We'll follow a fictitious example of a social experiment as we walk through some of the basic decisions about its design.

Using Pretests and Posttests to Measure the Effect of a Treatment

Let's say a startup company created a fancy new educational app, AddUpDog, which teaches kids math. We want to know if regular use of the app actually improves math skills. For our social experiment assessing the app's effectiveness, we decide that we'll find a group of 50 fourth-graders (our **research participants**, or *test subjects*) and then have them each use the math app for 10 hours. Then we will give them a standardized math test and see how they score.



Experimental designs are essential for testing causal relationships. One useful application of experiments is to evaluate the effectiveness of a specific program or intervention, such as whether the hypothetical math app described in this chapter raises the math test scores of the children who use it. *Elijah Kai Chen*

Let's visualize what we're doing. As we did in **Chapter 3: The Role of Theory in Research** with concept maps, it is helpful to create a *design diagram* for our experiment that lays out the key variables we're working with. In the design diagram illustrated in **Figure 12.1**, our independent variable is the use of the AddUpDog math app, and our dependent variable is children's scores on a standardized math test.

Even in the social sciences, scientists conducting experiments will call their independent variable the **stimulus** or **treatment**—language closely associated with lab experiments in the physical or medical sciences, where a specific intervention or factor is being observed within the lab setting. (The independent variable and dependent variable may also be referred to as the *presumed cause* and *presumed effect*, respectively—"presumed" because we do not yet have evidence that a change in our independent variable truly is the cause of any changes in our dependent variable.) In the study we're designing here (depicted in **Figure 12.1**), our stimulus is having our participants spend 10 hours on the math app. This happens during "Time 1," the start of the experiment. Then, during "Time 2" (say, a week after the participants use the app), we give them the math test.



Figure 12.1. One-Group Posttest-Only Design. At Time 1, the outset of the experiment, the treatment (X) is provided to the study participants (the experience of using the math app for 10 hours). At Time 2, a week later, we observe the students' scores on the math test. O1 refers to the first (and only) observation of the dependent variable, the test measuring students' math skills.

Let's say all the children score high on the test after using the app for 10 hours. Specifically, their average score was 95 out of 100. We happen to know that the average score for fourth-graders on this test is 73 out of 100. Clearly, the app works, right? The kids used the app, took the test, and scored very high. Ergo, it is empirically proven that this app makes kids smarter. Time to slap on a Science™ endorsement on AddUpDog's App Store description!

Not so fast. Has our experiment really given us strong evidence that AddUpDog improves test scores for fourth-graders? No, and that's because our experiment is not internally valid. One obvious problem with our experiment's design is that we don't have a previous test score (before using the app) from this group to use for comparison. In experimental lingo, the previous test would be called a **pretest**. As shown in **Figure 12.1**, in this experiment we only have the **posttest** scores (the scores *after* the stimulus is introduced). As a result, we can't really tell if the students improved or not. Yes, they scored high on our posttest, but it could be that we are experimenting on a class of fourth-graders who simply had very high math scores to begin with.

To address this problem, we decide to add a pretest (see **Figure 12.2**). Specifically, we give the kids the standardized math test during one given week (Time 1). We then have them do 10 hours on the app (Time 2). The following week (Time 3), we give them the same—or a similar—test to see if their scores have improved (we'll talk later about why or why not we might not give them the exact same test). A change in their scores between those time periods—positive or negative—would give us some sense that using the app had an effect on the kids' measured math skills.

Note, too, that by tracking the values of the dependent variable before and after, we establish temporal precedence, the second criterion for causality we discussed earlier. If the independent variable truly does have an impact on the dependent variable, introducing the stimulus will be followed by an increase in the value of the dependent variable sometime between the pretest and posttest. This outcome tells us something about the *direction* of any causal relationship that exists. If we had only cross-sectional observational data on students' app usage and math scores, for instance, we might observe a strong correlation between the two variables, but we would not be able to say if the app was raising math scores, or if higher-scoring math students happened to use the app more. In other words, the use of this experimental design addresses problems of **reverse causality** (described in **Chapter 3: The Role of Theory in Research**)—when what we think is the independent variable is actually the dependent variable.



Figure 12.2. One-Group Pre-Post Design. Here, we have two tests of the students' math skills: O1 (the first observation, the pretest) and O2 (the second observation, the posttest). If the students' test scores increase when we do the posttest, is our treatment (e.g., using the math app) responsible? We still can't tell, because several possible phenomena may be responsible for the increase: history, maturation, testing, instrumentation, and regression to the mean.

Our revised experiment has what we call a *pretest–posttest design* (or just *pre-post design* for short). With both tests in place, we can tell how much the students' test scores changed after using the app between Time 1 and Time 3. Let's say the average score jumped between the pretest and posttest—from 71 to 95. That's a 14-point increase! Obviously, our math app is responsible. We should report our results to the Department of Education, maybe fire off an email to the Gates Foundation about getting an education innovation grant . . .

Hmmm, or might there still be other reasons that the kids' scores went up?

Threats to Internal Validity When You Don't Have Control Groups

Even with the revised experimental design we've just implemented, we remain unable to rule out many possible alternative explanations for the results we observed in our experiment. Here are some important additional factors we need to consider, which fall under the broad category of *threats to internal validity*:

1. **History.** The change we observed over time in our dependent variable might have been caused in whole or in part by an extraneous event—something that happened *outside* our lab setting but that nevertheless affected the results we measured in the lab. We call this the **history threat** to internal validity (or *history effect*). This might indeed be a "historical" event, such as an event that affected all of society at the same time (say, a terrorist attack during the time period when we're doing an experiment relating to people's views on national security), but the term "history" here really refers to any event outside of the parameters of the experiment that nonetheless may be influencing our results—for example, a class that our participants also happen to be taking at the same time they're being studied. The important thing is that the outside event in question occurred sometime between our pretest and posttest, which means we may mistake the impact of that extraneous event for the impact of our experimental treatment. In our math app example, it is possible that in the time between our pretest and posttest, some of the students could have watched an educational television program that was the true cause of their increased scores. Or the students may have had a refreshing recess break right before their posttest (but not before the pretest), meaning they were better able to focus on their second go.
2. **Maturation.** Especially during a long-running study, we need to account for the possibility that any changes we

observe in the dependent variable between tests may have been caused by natural changes in our participants (a **maturation threat** or *maturation effect*) rather than by the experimental treatment per se. Say we are conducting an evaluation to see if a preschool program helps improve academic achievement later in childhood. The program—and therefore our evaluation of it—lasts for several years. By the time we conduct a posttest, the participants have become older. They have experienced general improvements in their intellectual ability that go beyond the impact of that one program (if it had an impact to begin with). Much like we saw with the history effect, lots of things may be going on simultaneously during the course of a long-term study like this one. Unless we could account for them, we would not want to chalk up any improvements in the children's cognitive abilities just to the influence of the preschool program.

3. **Testing.** The **testing threat** to internal validity (or *testing effect*) involves concerns that giving our test subjects the pretest could itself have influenced their posttest scores. In our example, just showing the fourth-graders our math test would make them more familiar with the test format and content, meaning they are likely to score higher on the posttest than they otherwise would have. This bias toward improved scores on any evaluation done multiple times is rather obvious: after all, this is why test preparation companies make millions of dollars getting students to take pricey courses or buy practice materials to give them an edge on the SAT, ACT, GRE, and other standardized tests.
4. **Instrumentation.** In our math app example, suppose we decide we don't want to give our fourth-graders the exact same math test as a posttest, given that their scores may increase because of the testing effect (i.e., because they are familiar with the test from the first time they took it). We change the questions on the posttest so that they're equally hard but not exactly the same. The problem, of course, is that we need to be sure our posttest actually is equally hard, so that any changes in students' scores can't be attributed to the changed nature of the test. This latter problem we're dealing with is called an instrumentation threat (or *instrumentation effect*)—the possibility that the difference between our pretest and posttest scores isn't due to the experimental stimulus or treatment, but rather to changes in the administered test (our study's "instrument"). If the posttest we give our fourth-graders has a lower degree of difficulty than the pretest, this instrumentation issue will give them, on average, higher posttest scores. In turn, we may conclude, erroneously, that the app had an effect. Conversely, if the pretest is easier than the posttest, the kids will have lower scores on the posttest, and our experimental results may understate the effectiveness of the app. The instrumentation threat also applies more broadly to cases in which the kind of testing we do itself shapes our results to an extent that we may mistake that effect for the influence of the independent variable.
5. **Regression to the mean.** Another possible explanation for our findings is a phenomenon called **regression to the mean** (also called just *regression*, but note that this is different from another type of "regression" we'll talk about in **Chapter 14: Quantitative Data Analysis**, which has to do with using statistics to measure the relationship between two variables). Regression to the mean refers to a statistical tendency that may make us observe changes in the dependent variable across the pretest and posttest that don't have to do with the influence of the independent variable. Specifically, we may have a situation in which a group's performance on the pretest happens to be particularly extreme, and their later results on that measure move toward what we typically expect—the mean—simply because the second test, the posttest, didn't happen to have such extreme cases. Let's say some of our fourth-graders using the math app are having a bad day and happen to score very low on our pretest. By statistical theory, they are likely to score higher on the posttest (closer to the mean or average). Therefore, their improved scores in this case will be due to regression to the mean, rather than to the influence of the math app. (Note that the exact opposite could also be true—the students happen to score high on the pretest, and then, because of regression, they score lower on the posttest.) Regression to the mean is often a problem when there is a cutoff for admission into an experiment. Say we only pick students for our study who score low on their math scores because we want to see whether the app is effective for struggling students. Because of this research design, however, we should expect some regression to the mean—that is, the results of the posttest to be generally higher. Some of the students selected for our study just had exceptionally bad luck on that first test, and their scores on subsequent tests will be closer to their average performance.

So, how do we account for all these possible alternative explanations in our experiment? The standard scientific approach is to add a **control group**. A control group is a sample of participants who do not receive the experimental stimulus. Nevertheless, we include them in our study and measure their scores on the dependent variable.

What does having a control group give us? As illustrated in **Figure 12.3**, the control group's outcomes can be compared with those of the **experimental group** (also known as the *treatment group*), which is composed of the participants who do receive the stimulus or treatment. Setting up the experiment in this way helps rule out the threats to internal validity we've just described. *Both* the control group and treatment group should be subject to the threats to internal validity mentioned earlier (e.g., history, testing, regression), meaning we still should be able to measure the effect of the stimulus, which only the experimental group received.



Figure 12.3. Two-Group Pretest–Posttest Design. A control group allows the researcher to isolate the effect of the treatment and measure the actual difference it makes—everything else being equal.

Threats to Internal Validity When You Don't Have Randomization



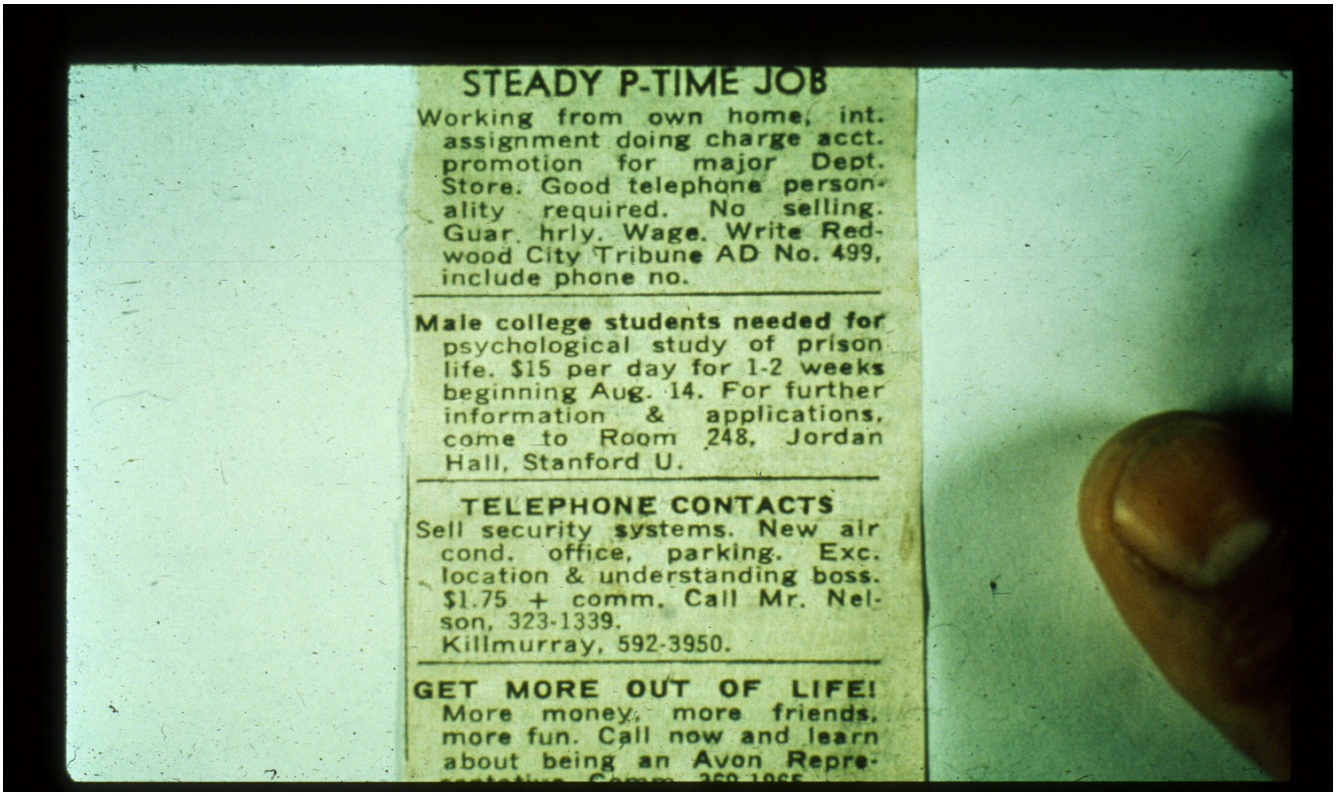
Threats to internal validity occur when we have experimental and control groups that differ on key variables because of selection bias. If the groups are different from the outset, then you are essentially "comparing apples to oranges." It will be unclear whether any differences observed after the experiment are due to the treatment or to other extraneous factors. *Anastasiya Romanova, via Unsplash*

Let's say that the average posttest score for our experimental group—the children who used the math app for 10 hours—does turn out to be higher than the average score for our control group who didn't use it. Can we finally conclude that the math app is effective?

Not yet, unfortunately. It's possible that the differences we observed between the control and experimental groups existed *before* we administered the stimulus to the experimental group. As a result, we may falsely believe that the independent variable explains the two group's divergent scores when preexisting differences really do. In our example, the fourth-graders in our experimental group may have been better at math than the control group students were even before they started using the math app. This is what we call a **selection threat** (also known as a *selection effect*). In this

scenario, a certain kind of participant is more likely to be selected for the experimental group, and as a result, the two groups diverge from one another on the outcome being measured in ways that have nothing to do with the stimulus given to the experimental group.

For instance, let's say that when we were looking for kids to test our AddUpDog app, we sent out a form to fourth-grade classes asking for volunteers. We asked the students' parents if they would be willing to have their child participate in the study for $20, and we gave them the option to be in the experimental group or control group. Many parents might have chosen the control group for their child because they wanted the money but didn't want to sit their kid down to use an educational app for 10 hours. Let's say that this was especially the case for the parents of kids who were already not doing well in math class. Then our control group might perform more poorly than the experimental group on our math test regardless of whether or not the app works. Students who struggle at math were just more likely to be selected for the control group to begin with. The differences between our two groups created bias—specifically, **selection bias**—in our results.



The newspaper advertisement used to recruit participants for the Stanford Prison Experiment, which appeared in the "help wanted" sections of the *Palo Alto Times* and the *Stanford Daily* in August 1972. The use of the words "prison life" may have created selection bias in the study's sample, drawing more aggressive individuals whose personalities accounted for the prisoner abuse observed in that study. *Stanford Prison Experiment, via PrisonExp.org*

You might be wondering why the pretest doesn't help us avoid this problem of selection. The answer is that it does: if we do a pretest, and the experimental and control groups are different on key variables, we clearly have a problem. But what do we do then? Even with the pretest alerting us to the fact that our comparison groups are not alike, we're still stuck with our selection problem. In our math app example, let's say we decide to have one fourth-grade class serve as our experimental group, and another fourth-grade class serve as our control group. Then, through our pretest, we discover

that the students in one of these classes happen to be math geniuses. At this point, do we really need to recruit an entirely new sample of participants? Maybe, but luckily we can avoid this problem altogether if we set up our experiment with *randomized* control and treatment groups, as we will describe in the next section.

Selection bias is a major problem in the social sciences, and it occurs more frequently than we might expect. For instance, in the Stanford Prison Experiment we discussed in **Chapter 8: Ethics**, one of the criticisms of the study's research design (beyond all the ethical problems we discussed earlier) was that the men who actually volunteered for the study were a special, self-selected group. The newspaper advertisement that recruited the study's participants mentioned that the researchers were conducting a "psychological study of prison life." A study by Thomas Carnahan and Sam McFarland (2007) used almost the exact same text for a newspaper ad recruiting study participants, except sometimes they excluded mention of "prison life" (thereby creating a control group in Carnahan and McFarland's experiment). They found that those who responded to the ad with the text "prison life" scored higher on measures of aggressiveness, authoritarianism, Machiavellianism, narcissism, and social dominance and lower on measures of empathy and altruism than responders to the ad that didn't mention "prison life." In other words, the people who responded to the Stanford Prison Experiment's ad may have been more sadistic to begin with, which may suggest that the guards' abusive behaviors during the course of that study may have been due at least in part to that selection bias rather than to the prison scenario, as the original study claimed.

Another issue related to the study sample is **attrition bias**. This refers to the possibility that certain types of participants may be more likely to drop out of our study. If the experimental and control groups are different in terms of *who* drops out of each, that difference may actually be responsible for our study's results, rather than the influence of the independent variable. (Sometimes this bias is called *mortality bias*, given how test subjects in a classic scientific lab experiment would die out or otherwise leave the study.) In our app example, if for some reason low-performing students disproportionately drop out of the experimental group (perhaps because the app bores them and they don't want to do the full 10 hours of screen time), we would end up only with high-performing students in that group. This could produce a higher posttest average in the experimental group than in the control group. The key problem is not so much that people dropped out, but that (1) they dropped out to a different extent across the two groups, and (2) those who dropped out had disproportionately high or low scores on the dependent variable. (We'll return to this issue of attrition bias and its close cousin, **nonresponse bias**, in **Chapter 13: Surveys**.)

How do we actually address the threats to internal validity posed by selection bias and attrition bias? We'll want to find a way to make sure that the experimental and control groups in our study start out as similar as possible on the variables that matter. We'll also want to make sure they remain that way, and that a disproportionate number of dropouts in one group is not skewing our observed outcomes. Randomly assigning participants to these two groups is the most straightforward way to address selection bias, as we will discuss in the next section. As for attrition bias, experimenters will go to great pains to ensure that members in both groups stay in the study, but there is often no easy solution to this problem when it crops up.

Key Takeaways

1. With a pre-experimental design, a study is vulnerable to five threats to its internal validity: history, maturation, testing, instrumentation, and regression to the mean.
2. Selection bias—when the experimental and control groups are different from one another at the beginning of an experiment—can also threaten the internal validity of a study. Random assignment to

control and experimental groups can address this bias because it ensures that the two groups are equivalent.

12.2. Randomized Controlled Designs: What You Should Do

Learning Objectives

1. Identify the defining characteristics of a classical experimental design that enable it to achieve internal validity.
2. Understand how to evaluate the results of a classical experimental design by comparing outcomes across its experimental and control groups.

What we call a "true" or **classical experimental design** is the gold standard for conducting experiments, a strategy that maximizes the internal validity of our study. As illustrated in Figure 12.4, a classical experimental design has all the features we've just described: a stimulus or treatment, an experimental and control group, and a pretest and posttest (although, as we'll discuss later, sometimes experimenters for various reasons choose not to conduct a pretest). But we need to go one step further to avoid the selection effect—the danger that our experimental and control group may start off differently on the key outcome measure we're interested in, thereby leading us to false conclusions about the effect of our treatment on the outcome measure. Specifically, we will want to assign participants into our two groups using a simple statistical technique that increases the likelihood that the experimental and control groups will be similar: **randomization**.

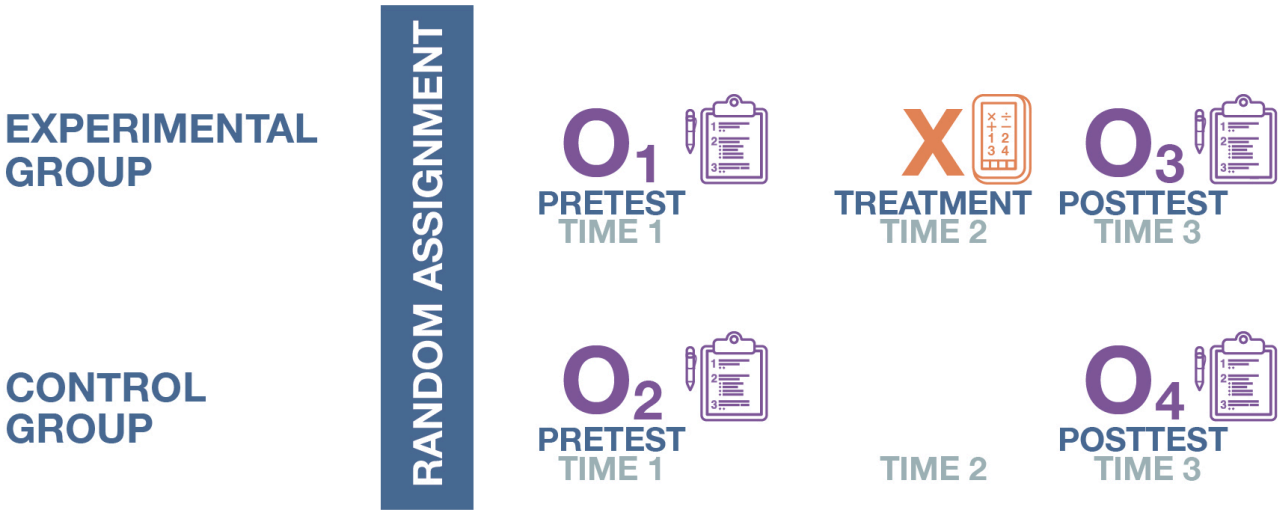


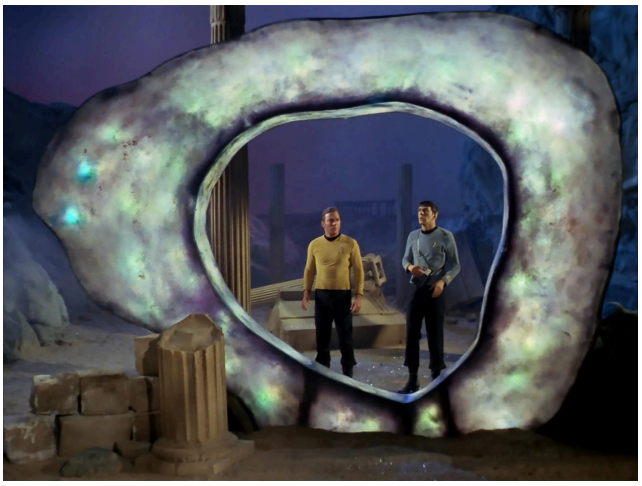
Figure 12.4. A Classic Experiment: Two-Group Randomized Pre-Post Design. This diagram illustrates randomization in a classical experimental design. Since every participant is assigned to the experimental or control group, every participant has the same 50 percent chance of receiving the stimulus or treatment. For added security, the pretests help us check the success of this randomization process.

Random selection is an important component of **probability sampling** (described in **Chapter 6: Sampling**). As you might remember, one method of making sure that your sample is truly representative of the population of interest is to gather a random sample—say, assigning a number to each member of the target population and then using a random number generator to choose your sample's cases from that enumerated list. In experimental design, the use of randomization is a bit different. Instead of selecting cases at random from a larger population, *random assignment* within an experiment means that we assign each of the study's participants to either the experimental and control groups at random—that is, with an equal 50 percent chance to be assigned to either group. As illustrated in **Figure 12.4**, if two students agree to participate in our math app experiment, we could flip a coin to assign the first student to the experimental group (heads) or the control group (tails), with the second student automatically assigned to the other group. We can continue doing this until we have a sufficient number of people in the two groups. (In practice, real researchers typically just use computer programs to assign their participants instantly to either group.)

What's the point of all this? Using randomization means that there is unlikely to be any substantial differences between the experimental and control groups at the outset of the experiment. By "any substantial differences," we mean that the two groups will obviously not be exactly the same on all characteristics, but they will usually be quite similar. After we randomly assign fourth-graders to the experimental and control groups in our math app example, the average math scores for the two groups are likely to be close on any pretest we give them. In fact, the two groups are likely to have similar characteristics on *any* measure—a similar proportion of brown-eyed participants, similar preferences for K-pop music, similar enthusiasm for pickleball—given that we placed everyone at random in our two groups. On occasion, we may come across extreme cases of participants who skew the measures for our otherwise neatly comparable experimental and control groups, but so long as we are faithfully using random assignment, the *average* characteristics should be more or less the same across the two groups.

What's more, simple statistical theory tells us that the *larger* the participant pool we're drawing on for our experimental and control groups, the *less likely* there are to be any meaningful divergences between the two groups. For example, if we're gathering participants for our AddUpDog evaluation from a pool of just 10 fourth-graders (therefore assigning five participants to each group), our randomized selection process may very well wind up stacking four or five math geniuses in one of our groups. Yet, once we reach a pool of, say, 100 students (50 participants in the experimental and control groups), the magic of randomization will have muted any differences in math ability between the two groups, making us confident that we can compare their outcomes in an apples-to-apples manner after we administer the AddUpDog treatment.

It may be helpful at this point to review the various benefits of a true experimental design for *causal inference* (again, our ability to conclude that a change in our independent variable truly causes a change in the dependent variable). By manipulating the independent variable and keeping all other factors constant within the lab setting, we are able to rule out any confounders that may independently influence the dependent variable. If we just examined the correlation between two variables using observational data, it would be quite easy to be tricked by spurious correlations, but through the use of control groups and random assignment, we have accounted for all alternative explanations for the outcomes we see. For this reason, drug trials and other clinical studies that use this classical structure are called **randomized controlled trials** (RCTs)—"randomized" referring to random assignment, and "controlled" referring to the use of treatment and control groups. Other variants of the term are *randomized controlled experiments* or *randomized controlled studies*.



In a true experiment, the randomly assigned control group allows us to understand the *counterfactual*: what would have happened to the experimental group if they had not received the treatment. In this sense, you might think of the control group as an alternative timeline (like the one depicted in the classic *Star Trek* episode "The City on the Edge of Forever"), splitting off when the treatment was given. The difference between the experimental group's reality and that alternate reality tells us what impact the treatment actually has. CBS/Paramount, via *Trek Report*

With this set of safeguards in place, a classical experimental design allows us to conclude with confidence that we have measured the impact of the independent variable "all other things being equal" (a condition known by the Latin term **ceteris paribus**). This is a basic condition for being able to make a causal argument that any treatment is driving changes in the dependent variable, because it means we have ruled out any competing causal stories. And to repeat, random assignment across the experimental and control groups helps ensure this condition, because it makes the two groups comparable on all characteristics. In fact, with randomization, we can say the control group is a reasonable representation of our **counterfactual**: what would have happened to the experimental group if the treatment had not actually occurred. The counterfactual is like the science fiction notion of an alternate reality—had we gone back in time and chosen not to give the stimulus, what would we have seen? This is something that we never really know with observational data, but with a proper experimental design, the control group is an apt stand-in for our counterfactual.

Interpreting the Results of a Randomized Controlled Experiment

Now that we've randomly assigned our participants into the two groups, how do we interpret the results of our posttest? Picking up our example again, there are several things we want to look at to make a rigorous case that the AddUpDog app actually has—or does not have—an effect on students' math scores. First, the average math scores for the experimental and control groups on the pretest should be similar. Since the two groups were randomly assigned, this should not be a problem, as we've just discussed. Nevertheless, comparing the groups in this way is often worth doing, because it may very well be the case that their pretest scores are quite different—perhaps we did something wrong in the randomization process, or perhaps we were just unlucky. If the two groups do diverge substantially, we may need to recruit more participants or pursue other means (beyond the scope of this textbook) to address this issue.

Once we're confident in the similarity of our experimental and control groups, we can proceed with our analysis

by seeing if three distinct patterns appear in our results. These patterns need to hold for us to conclude that the relationship between the independent and dependent variables we're observing is truly causal. The three conditions are captured by the following three questions. Note that all three questions assume we are testing whether the stimulus or treatment *increases* scores on our outcome measure. In the AddUpDog example, for instance, we're evaluating whether using the app raises kids' math scores. The questions should be reversed (e.g., "higher" swapped out for "lower") if we are testing whether the stimulus or treatment actually *reduces* measured levels of the dependent variable.

> **Condition 1: Within-group pre-to-posttest change. Is the experimental group's posttest score higher than its pretest score?** If the treatment (here, math app usage) truly has an effect on the dependent variable (here, math test scores), then the posttest average for the experimental group *should be higher* than its pretest average.
>
> **Condition 2: Between-group posttest difference. Is the experimental group's posttest score higher than the control group's posttest score?** If the treatment truly has an effect on the dependent variable, then the experimental group's posttest average *should be higher* than the control group's posttest average.
>
> **Condition 3: Between-group differences in pre-to-posttest change. If the control group experienced an increase in its average score between the pretest and posttest, is that increase smaller than any increase experienced by the experimental group between its pretest and posttest?** Note that this condition is relevant *only* when the control group's scores increase in its scores between the pretest and posttest. And observing that increase by itself does not rule out a causal relationship between the independent and dependent variable. After all, it's possible that the various alternative factors we mentioned earlier (e.g., testing effects) are at work here, too, which may mean that everyone's scores—those for the experimental group *and* the control group—go up between the pretest and posttest. That said, if the control group experiences such an increase between the pretest and posttest, that increase *should be smaller* than the increase experienced by the experimental group.

If the answer to any of these questions is "no," then our results indicate that there is no cause-effect relationship between the independent variable and dependent variable. The stimulus does not by itself cause a change in the dependent variable. But if the first two conditions hold (and the third one as well, if it is applicable), then we can conclude that the causal relationship exists—in our math app example, that the AddUpDog app truly does improve math skills. In fact, we can quantify the size of that effect by calculating the **average treatment effect** (ATE)—the change in the dependent variable that we expect, on average, after the stimulus or treatment is received. The ATE in a classical experimental design would be the difference between the experimental and control groups' average scores on the posttest.

In **Figure 12.5**, we can see what the results of our AddUpDog experiment might actually look like. The numbers in the diagram refer to each group's average on the math test (pretest or posttest) out of 100 possible points. As you can see, the pretests we conducted gave us a similar baseline for the experimental and control groups—74 and 75, respectively. Between the pretest and posttest, the experimental group's math score rose from 74 to 86—satisfying Condition 1 that the experimental group's posttest must be higher than its pretest. In line with Condition 2, we see that the experimental group's posttest average of 86 is higher than the control group's posttest average, which is 76. Finally, the control group's math score rose slightly over the course of the experiment—from 75 to 76—meaning that Condition 3 is also relevant. When comparing scores for the two groups between the pretest and posttest, we see that the control group's 1-point increase is smaller than the 12-point increase experienced by the experimental group across the same tests, satisfying Condition 3. All three conditions hold, and therefore, we have evidence to believe that 10 hours of using our math app does lead to improvements in the math skills of fourth-graders. (Note the very specific way we're phrasing our findings here, which we will talk about more later in regard to the external validity of our experimental results.) Furthermore, we conclude that the average treatment effect here is 10 points—that is, using the app, on average, boosts math scores by 10 points.

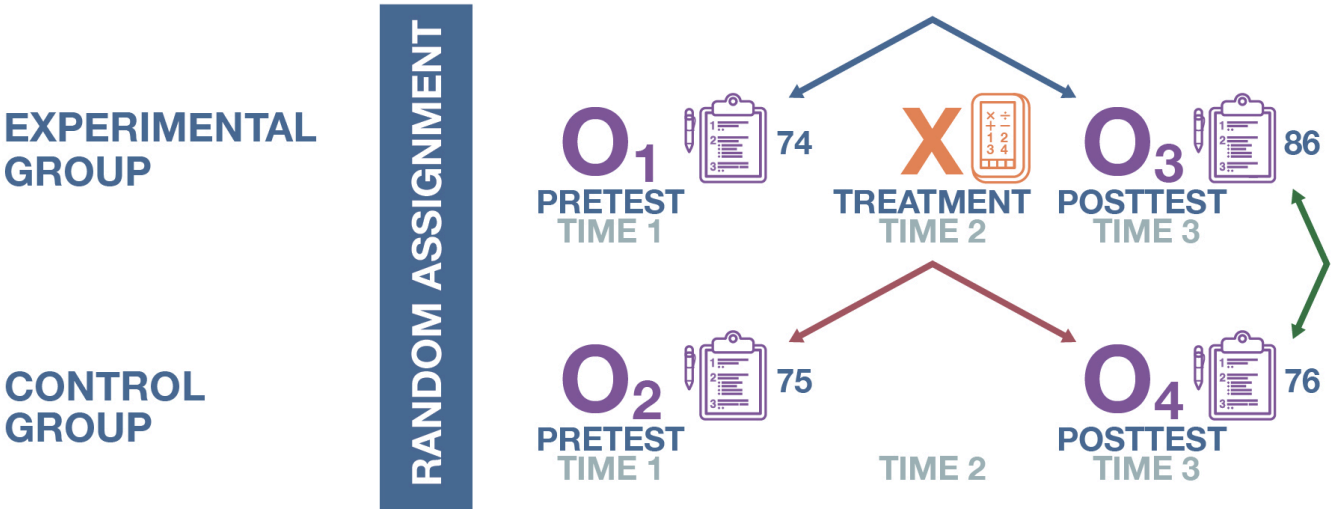


Figure 12.5. Two-Group Randomized Pre-Post Design with Scores (Math App Example). When we conduct a randomized controlled experiment, we should look for three patterns in our results. First, we should see if the experimental group's posttest score is *higher* than its pretest score (the comparison indicated with the blue arrows). Second, we should see if the experimental group's posttest score is *higher* than the control group's posttest score (the green arrows). Third, if the control group experienced an increase in its average score between the pretest and posttest (the red arrows), we should see if that increase is *smaller* than any increase experienced by the experimental group between its pretest and posttest (the blue arrows). If all three conditions hold—as they do in the previous example, based on the test scores indicated there—we can conclude that a change in the independent variable causes a change in the dependent variable.

Key Takeaways

1. A defining characteristic of a classical experiment is its random assignment of subjects to the control and experimental groups. This is a critical condition if the experiment is to have internal validity.
2. In a classical experiment, three outcomes would allow us to conclude that the treatment has an effect (in this example, a positive effect): (a) the experimental group's posttest is higher than the pretest; (b) the experimental group's posttest is higher than the control group's posttest; and (c) the difference between experimental group's posttest and pretest is larger than the difference for the control group.

12.3. Persistent Validity Problems: What You Still Need to Avoid

Learning Objectives

1. Discuss the similarities and differences between the major social interaction threats to internal validity.
2. Describe how a double-blind study design corrects for experimenter effects.
3. Explain why an experiment should include a placebo in situations in which participants have a belief in the treatment's efficacy.
4. Understand what external validity is and why it can be a problem for experiments.

To recap, a true experimental design includes both an experimental group that receives the stimulus and a control group that does not. As a result, the host of alternative factors that might potentially explain any change in our dependent variable—history, testing, instrumentation, maturation, regression—shouldn't matter. Those factors should affect the control and experimental groups in a similar fashion, and therefore we can rule out causal stories other than our hypothesis that our stimulus brought about the observed outcomes. In turn, randomly assigning participants to the experimental and control groups reduces the possibility that the two groups will be very different on any characteristic at the outset of the experiment, thus limiting the possibility of a selection effect. Incorporating these key checks within our experiment's design improves its internal validity and allows us to draw firmer conclusions about whether a causal relationship between the independent and dependent variables actually exists.

That said, a classical experimental design remains at risk of various social interaction threats to internal validity. We'll go through some of them next, and then talk about the larger vulnerability of experiments as a research method: its general lack of external validity. Please note that it is reasonable to expect any experimental study to have some problems of the sort described here, even if the experiment is rigorously designed. These limitations of internal and external validity should be acknowledged and discussed at length in the limitations section of your paper, but they do not necessarily condemn your study to irrelevance, and to some extent, these trade-offs are inevitable in the pursuit of any research strategy, including an experimental design.

Social Interaction Threats



A researcher's expectations can easily become self-fulfilling—a perplexing problem that falls under the umbrella of a larger psychological phenomenon known as the *Pygmalion effect* (named after the Greek myth of a sculptor whose adoration of a statue he carves makes the woman come to life). *Jean Raoux (*Pygmalion Adoring His Statue, *1717), via Wikimedia Commons*

Social research is a social activity. People—researchers and participants—interact with one another at every stage in the process. This opens a window for sources of potential bias to creep into our experiments. **Social interaction threats** are potential problems with internal validity that arise as researchers and study participants interact with each other over the course of an experiment (Trochim 2020). In these situations, social pressures may generate outcomes in the lab setting that are not actually caused by the treatment. Most of these threats occur because the various groups or key people involved in carrying out the research are aware of each other's existence and the role they play in the research project.

Here are the major social interaction threats to internal validity:

1. **Treatment diffusion and treatment imitation.** Both of these threats can occur when members of a control group learn about the treatment being provided within the study and adapt their behaviors in response to that knowledge. In our math app example, let's say the fourth-graders in the study's experimental and control groups came from the same school. The students using the app might share their experiences with it at lunch or on the playground. One possibility is that the children in the control group—blown away by AddUpDog's disruptive technology and canine-friendly interface—then decide to download the app on their own and use it avidly at home. As a result, there has been **treatment diffusion** across the two groups, even though the control group is not supposed to be exposed to the app. Another possibility is that the control-group kids get so excited about math apps in general that they stuff their phones and tablets with them and spend all their waking hours using them. In this case, members of the control group are engaging in **treatment imitation**—changing their math-learning behaviors based on their knowledge of the app-based tutoring the experimental group is receiving. Both of these examples of diffusion and imitation may affect the posttest performance of the control group, thereby jeopardizing our ability to calculate the direct impact that the math app has on students' test scores.
2. **Compensatory rivalry and resentful demoralization.** When **compensatory rivalry** is occurring, the control group knows what treatment the experimental group is getting and develops a competitive attitude with them. In our example, the fourth-graders in the control group might decide to compete with the experimental group "just to show them" how well they can do. In social contexts like these, participants may even be encouraged by well-meaning teachers or administrators to compete with each other. Although this might make educational sense as a motivation for the students in both groups to work harder, it works against the researcher's ability to determine whether the math app *by itself* has an effect on student performance. **Resentful demoralization** is basically the opposite of compensatory rivalry: here, members of the control group know what treatment the experimental group is receiving, but that knowledge doesn't encourage them to compete—instead, it discourages them. They get frustrated and angry and give up. Unlike the previous threats we've just discussed, resentful demoralization is likely to exaggerate rather than mute posttest differences between groups. In other words, if the fourth-graders in our control group got fed up with adding up numbers after learning they weren't selected to partake in the AddUpDog math experience, their poor scores on their posttest would make the app look even more effective than it actually is.
3. **Compensatory equalization of treatment.** The social interaction threat known as **compensatory equalization of treatment** directly involves a study's researchers as well as participants. When the control and experimental groups become aware of the conditions that the other group is experiencing, they may yearn to be in the other group. Often they will put pressure on the experimenters to have them reassigned to the other group. Alternatively, the researchers may feel obligated to compensate one group for the advantages that the other group is seen to be receiving. Either course of action can muddy the experiment's ability to accurately and precisely

determine the impact of the treatment on the dependent variable if it winds up sabotaging the randomization process, artificially making the experimental and control groups more similar to each other, or otherwise undermining the study's design.

4. **Experimenter expectancy.** Researchers may also bias their study's results if they expect the experimental and control groups to behave differently and then act in ways that further those expectations or make them known to participants. Let's say we are fervent believers in the AddUpDog app and therefore are convinced that the math scores of the fourth-graders in our experimental group will shoot up after they use the app. If so, we researchers may take actions—consciously or unconsciously—that influence their behaviors and possibly the study's findings as well. For example, we might unconsciously give the treatment group clearer instructions, more encouragement, or more time on the math exam they take for their posttest. That said, researchers don't necessarily need to take specific actions that help the treatment group to create this form of bias—we may simply give off cues to our participants that make them try harder or slack off in line with our expectations. For instance, if the fourth-graders in our treatment group learn that they are supposed to perform better than the control group, they may feel inspired to do so. That same information might demoralize the control-group kids, affecting their actual performance on test day.
   Note that this particular threat to internal validity arises first from the perceptions of the *researchers* (thus its name: **experimenter expectancy effect**). Somehow, these views filter down to the participants, who conform to researcher expectations. While you, as an ethical researcher, may swear up and down that you won't put your thumb on the scale, studies have shown just how much a difference an experimenter's expectations can make. Indeed, the experimenter expectancy effect is just one manifestation of a larger phenomenon called the *Pygmalion effect* (also known as the *Rosenthal effect*), which is described in **Video 12.1**. The psychologists Robert Rosenthal and Lenore Jacobson (1966) coined the term after observing how teacher expectations strongly influenced student performance in the classroom. (In Greek myth, Pygmalion falls so intensely in love with a sculpture of a woman he carves that she comes to life.) In their own experiment at an elementary school, Rosenthal and Jacobson took students in 18 classes over six grades and randomly assigned a fifth in each class to the experimental group, with the remainder serving as the control group. For each of the students in the treatment group, researchers told their teachers that they would show "unusual intellectual gains" over the year. Lo and behold, even though these students had been assigned at random to the treatment group, they ended up showing greater gains in measured IQ than the control group did, especially in the lower grades. In a similar fashion, researcher expectations that experiment participants will behave in a certain fashion can easily become self-fulfilling.



One or more interactive elements has been excluded from this version of the text. You can view them online here: https://viva.pressbooks.pub/sociology-research-methods/?p=407#oembed-1

Video 12.1. Experimenter Expectancy Effects. This video reviews some examples of how expectations of researchers can bias their studies in favor of confirmatory findings.

The social interaction threats we have described can be minimized. One strategy is to recruit our experimental and control groups in a way that does not make them aware of the other group's existence. For instance, we might pick our experimental group from one school, and our control group from another school (of course, this leaves our study open to the risks of selection bias we've discussed earlier). To avoid compensatory effects, we could make sure that everyone involved with the study knows how much the effectiveness of the experiment rests on preserving the original experimental and control groupings and not making any efforts to equalize outcomes across those groups. To deal with experimenter expectancy effects, we could conduct a **double-blind study**—not letting research staff who interact

with participants know who is in the experimental or control group, and not letting the participants themselves know whether they are in one group or another. Both experimenters and participants are thus "blinded" to perceptions that might influence their actions and change the outcomes of the experiment. (A similar approach is utilized in academic peer reviews: scholars are typically asked to review manuscripts without being told the names of the authors, and the authors are never told who their reviewers are, with the idea that this anonymity will prevent biases from influencing the conclusions of the reviewers.)

Table 12.1 summarizes all the threats to internal validity we have described in this chapter, along with possible solutions to them.

Table 12.1. Threats to Internal Validity and Ways to Address Them

Potential Bias	Issue	Solution
History effect	An occurrence outside the experimental setting influences the dependent variable, which could be mistaken for the direct effect of the stimulus or treatment.	Add a control group; both the experimental and control groups should experience any extraneous events, allowing researchers to see the impact of the independent variable alone on outcomes.
Maturation effect	Changes we observe in the dependent variable between tests may have been caused by natural changes in our participants, rather than by the experimental treatment per se.	Add a control group; both the experimental and control groups should change or mature over time similarly, allowing us to separate the treatment effect from the effect of the passage of time.
Testing effect	Giving our test subjects the pretest could influence their posttest scores.	Add a control group; if both groups take the same pretest, the overall results won't be affected. Using filler questions to hide the real purpose of the study can also help counter testing effects.
Selection bias	The experimental and control groups are different at the beginning of the study in ways that affect the study's results.	Randomly assign individuals to the control and experimental groups, which will make the two groups comparable.
Regression to the mean	Participants who score extremely high or extremely low on the pretest will tend to score closer to the middle (i.e., to the average or mean) on the next test, which may be mistaken for the impact of the treatment.	Randomly assign individuals to the control and experimental groups. Both groups should now be equally subject to regression.
Social Interaction Threats		
Treatment diffusion/ imitation	Members of the experimental and control groups may share their experiences, influencing outcomes for the control group. Control groups may imitate experimental treatments.	Prevent contact between the experimental and control groups as much as possible. One way to accomplish this is by recruiting our experimental and control groups in a way that does not make them aware of the other group's existence.
Compensatory rivalry	The control group knows what treatment the experimental group is getting and develops a competitive attitude with them.	Prevent contact between the experimental and control groups as much as possible. One way to accomplish this is by recruiting our experimental and control groups in a way that does not make them aware of the other group's existence.
Resentful demoralization	Members of the control group know what treatment the experimental group is receiving, and they become discouraged.	Prevent contact between the experimental and control groups as much as possible. One way to accomplish this is by recruiting our experimental and control groups in a way that does not make them aware of the other group's existence.

Compensatory equalization of treatment	Members of the control group learn about the conditions experienced by the experimental group and may put pressure on the experimenters to have them reassigned to the experimental group. Alternatively, the researchers may feel obligated to compensate one group for the advantages that the other group is seen to be receiving.	Prevent contact between the experimental and control groups as much as possible. One way to accomplish this is by recruiting our experimental and control groups in a way that does not make them aware of the other group's existence.
Experimenter expectancy effect (Pygmalion effect)	Researchers expect the experimental and control groups to behave differently and then act in ways that fulfill those expectations or make them known to participants.	Set up a double-blind study in which neither the experimenters nor the participants know who is in the experimental or control group.

Trick or Treatment: How Best to Manipulate the Independent Variable

In a groundbreaking social psychological experiment, Shelley Correll, Stephen Benard, and In Paik (2007) sought to understand what was truly driving the "motherhood penalty"—the observed trend of women's professional careers languishing after they have children. Social scientists had long debated two possible causes for this pause in career advancement and wage growth after women become mothers: (1) discrimination—that employers prefer nonmothers over mothers, believing that the latter won't be as productive or committed to the job—or (2) lifestyle choices—that mothers opt to prioritize family over career, leading them reduce their efforts in the workplace and avoid pursuing more demanding (and therefore higher-paying) jobs.



Shelley Correll, Stephen Benard, and In Paik (2007) conducted an experiment to test the hypothesis that mothers are discriminated against in the job market. Fake résumés and other job application materials designed to present two equivalently qualified candidates were given to participants, who were told to evaluate them for a job. The only difference between the two applications was that one set of materials subtly indicated that the candidate was a parent. In spite of the commensurate qualifications of the two candidates, the experimental group gave consistently lower ratings to mothers on perceived competence and recommended lower salaries. *Pavel Danilyuk, via Pexels*

Correll and her collaborators found a creative way to distinguish between these two possibilities and measure the impact of employer discrimination alone. In their lab, undergraduate participants were told that a California-based communications company was hiring a marketing director and wanted their feedback—as young and savvy media consumers—about whom to hire. They were individually presented with résumés and job application materials for two highly qualified women with comparable skills and experience. They were then asked to evaluate each candidate—scoring them on competence and commitment, deciding whether to recommend them for hire, and even choosing a salary offer within a set range.

Unbeknownst to the participants, the company was fictitious, and the (also fake) résumés were carefully constructed

to be equally qualified, yet not suspiciously so. The "mother" résumé subtly flagged the fact that the candidate was a mother with a reference to their volunteer role as a parent-teacher association (PTA) coordinator. In the "nonmother" résumé, that volunteering role was replaced by one not related to parenting—fundraising for a neighborhood association. Supplementary job application materials indicated that the "mother" candidate was married and had two children, while the "nonmother" candidate was described only as married. The researchers not only made the two fictitious job candidates appear equally qualified but also swapped who was the "mother" and who was the "nonmother" for half of their participants—thereby using the magic of random assignment to further ensure the treatment (mother) and control (nonmother) scenarios were equivalent.

In spite of the commensurate qualifications of the two candidates, their evaluations were startlingly at odds. Mothers received competence ratings that were substantially lower than those for nonmothers. They were also recommended lower salaries. The researchers used the same approach with fictitious résumés of men and found that fathers actually did *better* than nonfathers on some measures, such as their perceived level of commitment. Together, these findings provided powerful empirical support for the claim that mothers are actually discriminated against in the job market. After all, the fake résumés were designed to present equivalent candidates who were randomly assigned as "parent" or "nonparent" to the participants who reviewed them. Any differences in perceived competence and commitment across the candidates could thus be attributed to the fact that the participants knew whether each candidate was a parent and—consciously or not—had their ratings influenced by that knowledge.

The elegant research design of the "motherhood penalty" study exemplifies a number of best practices for experiments. For one thing, the tight control that laboratory researchers have over the experimental setting was exploited fully in this study. Rather than just having participants talk vaguely about the professional competence and commitment of parents and nonparents, the researchers used an elaborate cover story and put together meticulously constructed fake job materials to come as close as possible to the decision-making process that might occur in an actual hiring manager's office. (That said, we'll have more to say later about how the contrived nature of lab experiments in general is a key weakness of this method.) The researchers also included a range of different measures of worker competence and commitment—from getting participants to explicitly rate those qualities on numerical scales, to asking them to provide hiring recommendations and salary offers, to even having them decide how many days of lateness would be tolerated before candidates were no longer recommended for hire (mothers were granted less time than nonmothers on that measure, too).

Another aspect of the study's design that deserves praise is the sophisticated way the researchers manipulated their independent variable. The two templates of job materials were not only matched—that is, written to make the two candidates appear equally qualified—but each participant also had an equal chance of receiving materials where Template A was the parent (and Template B the nonparent) or Template A was the nonparent (and Template B the parent). This combination of matching the parent and nonparent candidates and randomly mixing up their job materials meant that the only real difference between the two candidates was parental status. As a result, the researchers could confidently rule out any explanations for the motherhood penalty other than people's stereotypes about working moms. Furthermore, by carefully and discreetly signaling parental status, the study ensured that what experimenters expected to find—that parenthood would matter in these hiring decisions—wasn't overly obvious to participants and didn't unduly influence their ratings.

Note, too, how the researchers essentially used a **placebo** in their experiment. As you probably know, randomized controlled trials to test the efficacy of drugs or other medical treatments will not simply give the experimental group the treatment and the control group nothing. A clinical drug trial will instead have the control group receive a sugar pill—ineffective and harmless—so that neither group knows whether they are receiving the actual drug. As doctors have long known, believing you are receiving an effective treatment will improve your health outcomes regardless of whether the treatment really works (again, expectations matter!). Using a sugar pill or its relevant equivalent avoids this **placebo**

effect, a benefit not due to the treatment but rather the recipient's belief in that treatment.[1] By offering this "trick" or "treatment" pill at random, clinical researchers can ensure that they are isolating the effect of the intervention and not capturing any other factors entangled with it.

For similar reasons, social science researchers need to ensure that their experimental and control groups have the exact same experience within the lab except for the presence or absence of the stimulus being studied. In the "motherhood penalty" study, the researchers didn't say that the job candidate was a PTA coordinator on the "parent" résumé while leaving that space blank on the "nonparent" résumé. Instead, they created another volunteer role for the nonparent—fundraiser for a neighborhood association. Had they not added the fundraiser role, a skeptic might wonder if the lower scores that mothers received was due to their additional volunteering responsibility—which the nonmothers did not have, and which employers might see as conflicting with their work obligation. In other words, not being meticulously precise in their operationalization of the treatment would have meant the researchers were not really testing the effect of parental status, but rather parental status and something else.



If an intervention is given to an experimental group, the mere act of administering something to that group may trigger a response. This is known as the *placebo effect*, and it is the reason that drug trials will have a control group take sugar pills while the experimental group receives the actual drug: doing so allows researchers to rule out a placebo effect because the control group also received something. In social scientific experiments, the equivalent of administering a placebo would be to do some sort of intervention other than the one being tested. For instance, in the "motherhood penalty" study, the "mother" job candidate's résumé signaled her parental status by noting that she volunteered for a PTA; the "nonmother" candidate's résumé also mentioned a volunteering role, but one unrelated to parenting. *Karolina Grabowska, via Pexels*

These details might seem minor, but they are the sorts of considerations that good researchers fret about when designing their experiments. Setting up the ideal lab environment to study a social phenomenon and effectively manipulating the independent variable within that setting require a great deal of precision and creativity. And even when you think you have gotten those things right, whether your lab results hold up in the real world is a more intractable problem, as we will discuss next.

1. In fact, a large body of research has accumulated that finds that even when individuals know they are taking a sugar pill, these so-called open-label placebos can still be more effective than no treatment whatsoever, which scientists believe may be due to the subjective effects of the medication-taking process (Charlesworth et al. 2017).

Problems with External Validity: When Your Results Do Not Generalize



Laboratory experiments have given us amazing insight into how people's thoughts, feelings, and behaviors are shaped by norms and other social factors, but a key weakness of this approach is that they take place in artificial and contrived environments, which may or may not be good proxies for what might occur in the real social world. *Michael Burrows, via Pexels*

As we noted in the introduction to this chapter, external validity is the Achilles' heel of the experimental method—a major vulnerability in an otherwise remarkably rigorous approach to social science. To expand on our previous definitions, *external validity* (also known as *generalizability*) refers to whether we can reasonably say that the results we observed in our study's sample would also hold up in the target population. More broadly, we might wonder if we could generalize our results to another population entirely (say, Canada if our sample came from the United States), or another social context (a Muslim-majority culture if we studied a Christian-majority culture), or another time period (today if our data is from 20 years ago).

Laboratory experiments are criticized for their contrived conditions. In many psychology experiments, the participants come to a classroom or lab to fill out a series of questionnaires or perform a carefully designed computerized task, perhaps after experiencing a mock social interaction or watching a video of one. Yes, the researchers may try their best to emulate a real-world social situation, but a lab experiment is by definition artificial and only an approximation of actual life—and possibly a pretty poor one at that. The participant's observations or actions in the lab setting may not reflect what would go on in the real world with real people interacting and real reputations or resources on the line. For example, Barbara Fredrickson and her colleagues (1998) conducted an experiment to study whether self-objectification produced body shaming, promoted unrestrained eating, and diminished math performance. They recruited undergraduate students to come to their lab on campus. Participants tried on a swimsuit or sweater alone in front of a full-length mirror. They then completed questionnaires on their attitudes toward body shaming and took a food taste test and a math test.

Does wearing a swimsuit before a mirror in a lab capture what it is like to feel shame about your body in real life? Would you actually experience some sort of body-shaming incident right before eating a meal or taking a test? Maybe, but the social scenario fabricated in this experiment is extreme in many ways, as it is in most lab studies. Their contrived situations can give us vital clues about how people think and act in their everyday lives, but they should also give us pause. For one thing, researchers need to be very clear about what can actually be generalized to the target population based on their experiment. They need to be aware of what we have previously described as a study's **scope conditions**, the situations in which a study's findings can reasonably be thought to apply. For instance, our AddUpDog evaluation may have shown a causal effect, but under what conditions does that effect hold? We studied 10 hours of app usage over a week, and then tested math skills. Would fewer hours also have an effect? How long does any effect last? Would a

different kind of math test give us different results? Would the app also work if we tested it on fourth-graders in another state or country? We don't know the answers to these questions, given the narrow context and treatment involved in our single study—and lab experiments can be particularly narrow in the contexts and treatments they test. In any case, we as researchers need to be upfront about these potential limits to a study's external validity.

It is also important to mention that the ability of experimenters to manipulate their independent variable in a valid and generalizable fashion often depends on subterfuge. In the "motherhood penalty" study, for example, participants were told they were providing feedback to a real startup, even though that was just a cover story. They were deceived in this way so that they would take their task of vetting the job candidates more seriously, better mimicking conditions in the real world. Some experiments will go even further, manipulating the conditions that participants experience by using **confederates**, individuals hired by the researchers (or sometimes the researchers themselves) who act in certain ways to maintain whatever cover story the study is using to test its treatment or stimulus. The use of confederates and deception more generally not only raises ethical issues of the sort we discussed in **Chapter 8: Ethics**, but also runs the risk that participants see through the subterfuge, potentially undermining the study. In fact, a couple participants in the "motherhood penalty" study gave responses to researchers that raised concerns they might be suspicious of the cover story being used, prompting the researchers to remove them from the study to avoid social interaction threats of the kinds we discussed earlier.

Another issue of external validity that bedevils lab experiments is their tendency to draw on a homogeneous population for their participants: specifically, college students from rich Western countries with major research universities. (We talked in **Chapter 6: Sampling** about how problematic these **WEIRD samples** can be.) College professors often have a subject pool available of students who must participate in a certain number of studies to meet a course requirement. Students can also be easily recruited on campus by posting descriptions of research studies—using in-person or online appeals—and offering a small amount of compensation. But the heavy reliance on this select group of experimental guinea pigs raises questions about whether the results of many lab experiments are broadly generalizable. By contrast, survey researchers usually recruit respondents directly from their populations of interest. They can easily defend the external validity of their research based on the mathematical principles of probability sampling.

To recap, laboratory experiments often run into problems with external validity because of (1) the contrived nature of the treatment and social context that researchers are trying to approximate within the lab setting; and (2) the use of homogeneous samples that may give us a false or skewed picture of social life in the real world. To deal with one or both of these concerns, some researchers pursue field experiments as a sort of middle ground—an experiment conducted in the real world, often with larger and more representative samples. We'll describe these types of experiments in the next section.

Deeper Dive: Problems with Pretests

We mentioned before that experimenters sometimes opt to drop the pretest and just conduct the posttest. After all, so long as they randomly assign their participants to the control and experimental groups and have sufficient numbers in both, they can reasonably expect the two groups to be fairly comparable. One benefit of this approach is that it saves resources that would have been spent pretesting both groups. Another is that the researcher avoids another potential problem of external validity—that is, the **interaction of testing and the experimental treatment** (yes, it's a mouthful).

If we use an experimental design with both a pretest and a posttest, we run the risk that any treatment effect

we observe may not be just because of the influence of our independent variable, but rather because of the combination of the stimulus *and* the pretest given to the treatment group. Confused? Let's make things more concrete with an example. Consider a social psychological experiment by Shannon K. McCoy and Brenda Major (2003) that sought to understand how perceptions of prejudice influence feelings of depression. All the study's participants were given a pretest to assess their levels of depression, which was the researchers' dependent variable. Participants then read an article suggesting that prejudice against a particular racial group was severe and pervasive. For participants randomly assigned to the experimental group, the racial group in question was their own racial group; for those in the control group, it was a group other than their own. As it turned out, the experimental group—those who read about pervasive prejudice against their own racial group—reported greater levels of depression than those in the control group.

What's the problem here? We might be worried that giving participants a pretest to gauge their level of depression and then having them read about prejudice against their own ethnic group might tip them off to the relationship between prejudice and depression being studied. That, in turn, may make the information about racism that the participant just ingested more salient, leading them to score higher on the depression posttest. As a result, the experiment would not be testing *just* the impact of learning about prejudice against one's racial group on one's measured levels of depression. It would be testing the impact of learning about that prejudice *and getting a depression test* on one's measured levels of depression.

This means we have an external validity issue: we cannot generalize, based on the results from this study, that reading about racism leads to greater depression, even though that's really the causal relationship we wanted to assess. In the real world, people aren't getting depression tests before they happen to read about racism. Therefore, the hit to their mental health that the treatment group suffered may not occur for them. The distinction may seem nitpicky, but it is implicit in the design of this study. To avoid this issue—the *interaction of testing and the experimental treatment*—we should think carefully about whether pretesting our participants will give them information that will affect their performance on the subsequent posttest. If so, the external validity of our study may be harmed. In that case, we may want to dispense with pretests (as illustrated in **Figure 12.6**) and instead rely on the power of randomization to ensure that our two groups are similar at the outset in terms of any characteristic that might matter.



Figure 12.6. Two-Group Randomized Posttest-Only Design. With this experimental design, we can no longer examine the differences between pretests and posttests for the experimental and control groups, but we can still obtain some evidence of the treatment's causal effect—if we see that the posttest for the experimental group is higher than the posttest for the control group (the comparison indicated with the green arrows).

Key Takeaways

1. Social interaction threats to internal validity include treatment diffusion, treatment imitation, compensatory rivalry, resentful demoralization, compensatory equalization of treatment, and experimenter expectancy effects.
2. A double-blind study design corrects for experimenter effects because neither the experimenter nor the participant knows whether the participant is in the experimental or control group.
3. If an intervention is given to an experimental group, the mere act of administering something to that

group may trigger a response (a placebo effect). To rule out this effect, social scientific studies will have the control group undergo some sort of intervention other than the one being tested on the experimental group.

4. Lab experiments cause issues with external validity because of (a) the contrived nature of the treatment and social context that researchers are trying to approximate within the lab setting; and (b) the use of homogeneous samples that may give us a false or skewed picture of social life in the real world.

12.4. Field Experiments: External Validity at the Cost of Control



With the cooperation of the Minneapolis police department, Lawrence Sherman and Richard Berk (1984) conducted a field experiment to test two competing theories about whether criminal punishment deters domestic violence. As part of the study, police officers were randomly assigned to one of three responses to domestic violence calls: arresting the accused batterer, forcing the accused to leave the residence for eight hours, or offering mediation. As is often the case with field experiments, however, the researchers had little control over their research site, and police officers frequently deviated from the experiment's procedures. *Kindel Media, via Pexels*

Learning Objectives

1. Describe the key characteristics of a field experiment.
2. Describe the methodological approach of an audit study.

A key strength of laboratory experiments is the high level of control that researchers have over what goes on in the study. Typically, they can direct the recruitment of participants, manipulate the independent variable directly, and decide on when and how to measure the dependent variables. Yet, as we discussed earlier, laboratory experiments are often conducted under conditions that seem artificial. The social processes uncovered under such conditions may not be generalizable to life outside the lab.

Field experiments are conducted in the field—that is, in an actual organization's office or an actual city's labor market. They are often used to gauge the effectiveness of a social intervention, such as a health, social welfare, or educational program. Because field experiments gather their data from real-world settings, they have higher external validity than their laboratory counterparts. But this strength of field experiments is also their weakness, because real-world settings cannot be controlled as easily as lab settings can be. And because field experiments tend to be costly to conduct and may even require the cooperation of outside parties like government agencies, researchers without a lot of funding and other external support are particularly constrained in what they can do.

For one thing, conducting a field experiment often makes it hard or impossible for the researcher to randomly assign participants into experimental and control groups. Emily Becher and her colleagues (2018) conducted a field experiment to evaluate Parents Forever, a University of Minnesota divorce education program developed in the 1990s to improve parental well-being, coparenting, and parent–child relationships following a divorce. However, the researchers did not have access to a randomly assigned control group. The divorcing parents participating in Parents Forever had been court-mandated to attend a two-hour divorce education program. If they were going to be the experimental group—providing evidence of Parents Forever's effectiveness—then the researchers would need to compare their outcomes with those from a sample of similar parents also going through divorce but not in that program. To construct this control group, the researchers recruited divorcing parents through an online platform providing workers for hire. As the researchers acknowledged, this approach was not ideal, and in their study, they referred to the group recruited online as a "comparison group" rather than a "control group" to recognize how it fell short of the classical experimental ideal. (We'll discuss comparison groups and the quasi-experimental approach of studies like this one in greater detail later.) Although the researchers ended up with two groups of divorcing parents to compare, the demographics of the two groups were notably different. Specifically, the comparison group was more racially diverse and more educated than the participants recruited from the Parents Forever program.

All of the study's participants completed a presurvey and a follow-up survey three months later, which found that the Parents Forever program led to improvements in parenting practices, adult quality of life, self-efficacy, and child conduct problems and peer problems. These results, however, may have been skewed by the fact that the field experimenters were not able to recruit randomly selected control and experimental groups. Among other things, the higher degree of diversity and education in the comparison group might have accounted for some of the observed outcomes that the analysis credited to the program.

Even when it is possible to use random assignment to experimental and control groups in a field experiment, the lack of control that researchers have over conditions on the ground means that their best-laid plans often go awry. For example, researchers Lawrence Sherman and Richard Berk (1984) conducted a field experiment to test two competing theories about whether criminal punishment deters domestic violence. According to deterrence theory, punishment should reduce deviance—and thus arresting an accused spouse batterer should reduce future incidents of violence. Labeling theory, however, predicts the opposite, arguing that the act of arresting the alleged perpetrator will only heighten their self-identity as a batterer and therefore lead to more violence.

Remarkably, the researchers were able to convince the Minneapolis police department to help them determine which theory held up—that is, whether arresting alleged perpetrators actually reduced subsequent reports of domestic violence. The police department agreed to implement a policy of randomly choosing among three responses to domestic violence calls: arresting the accused batterer, ordering the accused to leave the residence for eight hours (separation),

or offering mediation (advice).[1] By introducing this randomization, the researchers essentially created experimental and (two) control groups, allowing them to see whether arrests were more or less effective than the other two responses in reducing future violence. This analysis found that arrested suspects were less likely to commit further acts of violence over the next six months than those who were ordered to leave (according to official data) and those who were advised (according to victim reports).

Unfortunately, the randomization was not as successful as the researchers had hoped. Police officers did not reliably follow the protocol they had been given. They would leave the study instructions at the station. They would misunderstand or ignore what they were supposed to do. Officers were much more likely to follow through on their randomly chosen directives when it involved arresting the accused perpetrator: they did so 99 percent of the time, but only 78 percent of the time when instructed to provide advice, and only 73 percent of the time when instructed to separate the alleged batterer from the residence. The implementation problems that the researchers experienced meant that their sophisticated research design failed to perform as well as it could have, hampering their ability to rigorously assess whether arrests did or did not lead to more incidents of domestic violence. (Essentially, the researchers were dealing with attrition bias: cases in their sample were dropped disproportionately from the control groups, meaning that the effect of *not* arresting alleged batterers could not be accurately measured.) For instance, it may have been that when officers encountered cases of spousal abuse that they perceived as especially severe, they were more likely to ditch any instructions to the contrary and go ahead and arrest the accused batterer. These severe cases would be more likely to be followed by further incidents of domestic abuse, and yet they were disproportionately dropped from the control groups of separation and advice, thereby overstating the effectiveness of such non-arrest responses in reducing future violence.

The Minneapolis domestic violence study underscores how field experimenters have less control over the conditions of the experiment than those working in laboratories. With enough resources and institutional support, however, it is possible to create field experiments that adopt all aspects of classical experimental design and minimize threats to internal variety in a robust fashion. Consider the Moving to Opportunity (MTO) study, a celebrated field experiment conducted in the 1990s that examined whether having residents of poor neighborhoods move to higher-income areas improved various life outcomes, especially for children. Sponsored by the U.S. Department of Housing and Urban Development (HUD), the study recruited 4,600 low-income families with children who were living in public housing projects. Even though the study was a field experiment, the researchers were able to implement a randomized controlled experimental design. They used a lottery to assign the families who volunteered for the study into one of three groups. One group received housing vouchers that for the first year could be used only to move to low-poverty neighborhoods. They also received counseling to help them find homes in these neighborhoods, and after a year, they could use their vouchers to move anywhere. The second group received housing vouchers without any geographic restrictions, but they did not receive any counseling. A third group did not receive vouchers but remained eligible for other forms of government assistance.

Early results were mixed, but a later analysis of the study's longitudinal data (depicted in **Figure 12.7**) by the economists Raj Chetty, Nathaniel Hendren, and Lawrence Katz (2016) found that those children who moved at a young age to low-poverty neighborhoods had later earnings that were, on average, a third higher than what children who didn't move wound up making as adults. These children were also more likely to go to college and grow up with a father present. (The same benefits did not accrue to children who moved when they were 13 years old or older, which helped explain some of the study's previous, mixed results: whether the vouchers made a difference for children appeared to depend

1. This protocol was followed only for less serious calls, the researchers note: "Cases of life-threatening or severe injury, usually labeled as a felony (aggravated assault), were excluded from the design for ethical reasons" (Sherman and Berk 1984:263).

on whether they moved early enough to have meaningful exposure to a low-poverty neighborhood during their most formative years.) Children in families that received unrestricted housing vouchers also saw improvements over non-movers, yet not to the same extent as for that first group. These experimental findings provided strong causal evidence in support of the view that the neighborhood where a child grows up decisively shapes their later life outcomes (Wolfers 2015).

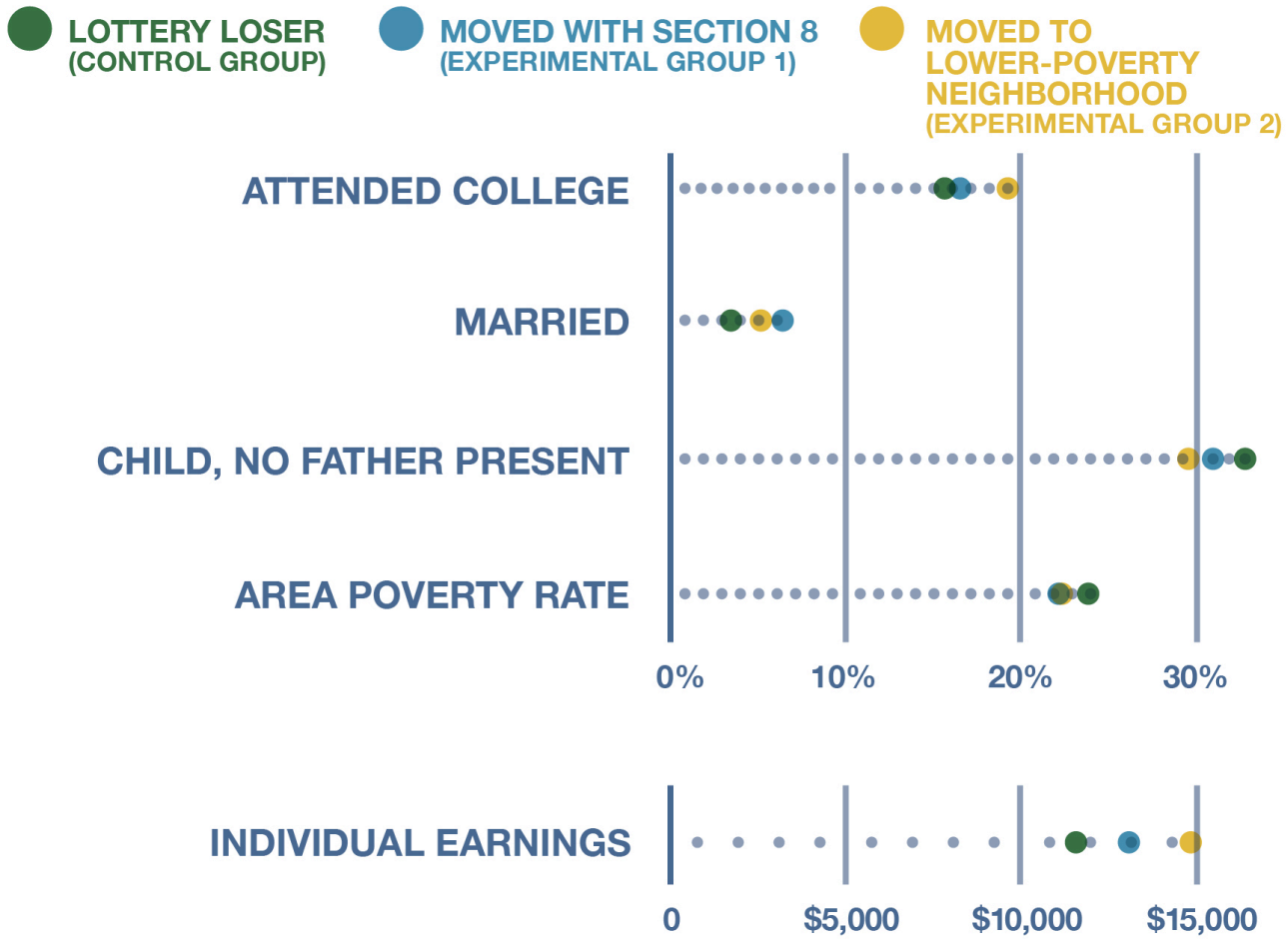


Figure 12.7. Outcomes at 24 Years or Older for Children Who Moved Before They Were 13: An Analysis of Moving to Opportunity Data by Chetty et al. (2016). The MTO field experiment examined whether residents of poor neighborhoods who moved to higher-income areas fared better in life. As part of the study, low-income families with children who were living in public housing took part in a lottery that assigned them to one of three groups. Children in the two experimental groups—the families who received housing vouchers that allowed them to move elsewhere—did better on various measures later in life, with the effects especially pronounced for those whose vouchers had only allowed them to move to low-poverty neighborhoods. *Graphic adapted from a* New York Times *graphic and updated using figures from Chetty et al. (2016)*

An important aspect of field experiments like MTO that evaluate policies is they do not just give the treatment (e.g., the housing voucher, the income support) to everyone. Instead, they have a lottery, with the lottery winners winding up in the experimental group (who receive the program's benefits) and the lottery losers serving as a control group. This enables researchers to rule out alternative factors and precisely measure the impact of the policy itself. Because the evaluation used for the Parents Forever divorce education program (described earlier) did not have such a lottery for entry into the program, it was much harder for researchers to say that any improvements they saw among the program's

participants were due to the program and not preexisting differences between that group and the cobbled-together comparison group.

Audit Studies: A Powerful Test of Discrimination

One frequently used type of field experiment is the **audit study**. Social scientists typically employ audit studies to test discrimination in the labor market, the real estate market, and other settings. The researchers will hire and train **testers** (also known as *auditors*) who are matched on all characteristics except one—for instance, gender or race—so that the study can see if the testers are treated differently based on that sole characteristic. The testers then go undercover into the particular setting being studied, strictly adhering to the behaviors modeled for them by researchers so that any bias they experience can be attributed solely to discrimination. This covert experimental approach is a very effective way to understand how discrimination operates in the real world, especially in comparison to methods like in-depth interviews and surveys, where respondents have strong incentives to hide racist or other socially undesirable attitudes.



In addition to their research applications, audit studies are frequently used by government agencies to test racial discrimination in housing markets. Auditors will pose as homebuyers or renters and see if real estate agents and landlords discriminate when showing them listings. *Thirdman, via Pexels*

'Fair housing groups and activists pioneered housing audits in the 1940s, and government agencies like HUD eventually adopted the strategy, conducting in-person audits to gauge levels of discrimination in rental and housing markets (Gaddis and DiRago 2023). Think tanks like the Urban Institute later applied the approach to studying discrimination in hiring. Over the past several decades, the shift toward applying online for jobs and housing has made it easier for researchers to conduct large-scale audit studies—by electronically disseminating materials for fictitious job candidates, for instance, rather than training closely matched auditors and sending them into the field.

The audit methodology has also been extended well beyond its original domains of housing and job markets. Heather Kugelmass (2016) used it to study whether mental health therapists discriminated against certain types of people when accepting new patients. She hired a group of voice-over professionals as auditors and gave them a script for requesting a therapy appointment over the phone. These auditors adopted racially distinctive names and used speech patterns that signaled their racial and class backgrounds. They then called 640 psychotherapists randomly drawn from a private insurer's directory and left messages asking for help with their depression and anxiety.

As the audit methodology requires, the study's help-seekers presented themselves in the same way across every relevant characteristic and behavior except those relating to their class, gender, and race. All of them named the same health insurance plan. All of them requested a weekday evening appointment. And all of them asked the therapists to leave them a voicemail message with any available appointment times. As it turned out, therapists were much more likely to offer an appointment to white middle-class help-seekers (see **Figure 12.8**). The gender of the help-seeker didn't appear to affect whether therapists returned their calls, but women did get more appointment offers for the preferred hours.

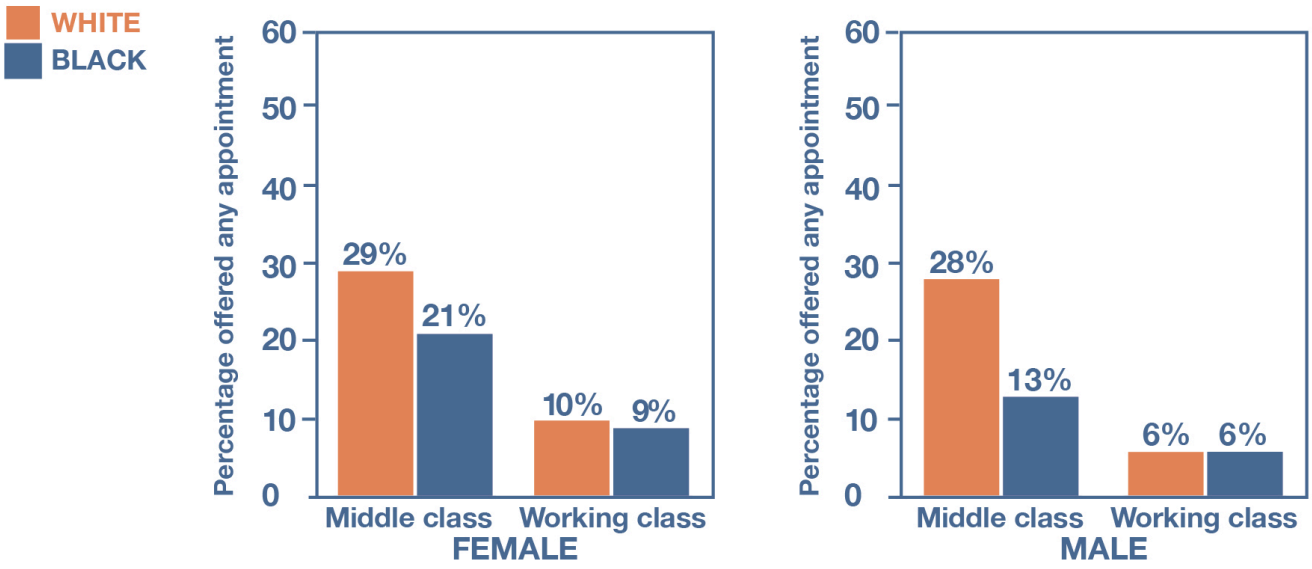


Figure 12.8. Appointments Obtained for Mental Health Therapy by the Class, Race, and Gender of the Help-Seeker: An Audit Study by Kugelmass (2016). Heather Kugelmass (2016) conducted an audit study examining whether mental health professionals discriminated against certain types of people when accepting new patients. She hired actors to impersonate help-seekers, having them use race- and class-specific ways of speaking when they called for appointments. The study found sizable class, racial, and gender biases—most strikingly, callers who signaled themselves as being from a middle-class background were much more likely to obtain an appointment. *Figure adapted from a chart in Kugelmass (2016)*

A series of influential studies using the audit methodology were conducted by the late sociologist Devah Pager and her collaborators starting in the early 2000s. For her *American Journal of Sociology* article "The Mark of a Criminal Record" (2003), Pager hired four young men to pretend to be job applicants for a wide range of entry-level jobs. Every week, they applied to 15 openings advertised in local job listings, auditing 350 employers in total by the end of the study. The four auditors—"two blacks and two whites"—were all college students who were 23 years old and "matched on the basis of physical appearance and general style of self-presentation" (946–47). Characteristics of their actual educational attainment and work experience were tweaked so that they were similar for the purposes of applying for jobs. Furthermore, among the two testers of each race, one was randomly assigned a "criminal record" for the first week: a felony drug conviction and 18 months of served prison time, which they would dutifully report to employers. (Compare this approach to the random assignment of parental status in the "motherhood penalty" study.)

The pair of testers then applied to the same jobs, one a day after the other, with the sequence randomly determined. Testers would visit the employer, fill out applications, and then see if they received a positive response, or "callback"—which could take the form of either a job offer on the spot or a phone call later on from the employer on a voicemail number set up for each tester. Every week, the criminal record was rotated among the racially matched pair.

What was the point of this matched and randomized setup? Pager wanted to isolate just the effect of race and just the effect of a criminal record on whether employers would hire a job candidate. With the treatment of the "criminal record" randomly applied to the pair of racially matched testers and changed every week, Pager could plausibly chalk up any differences in hiring outcomes directly to the criminal record, rather than to anything she had missed in matching the testers on every relevant characteristic.

As shown in **Figure 12.9**, Pager's audit study found that applicants of either race who had a criminal record were much less likely to receive callbacks. Perhaps what was most shocking, however, was that the African American testers

received so few callbacks. In fact, while only 17 percent of the applications by white testers with criminal records resulted in callbacks, only 14 percent of those by African American testers *with clean records* did so. "This suggests that being black in America today is essentially like having a felony conviction in terms of one's chances of finding employment," Pager said in a 2016 interview (Stanford Center on Poverty and Inequality 2016).

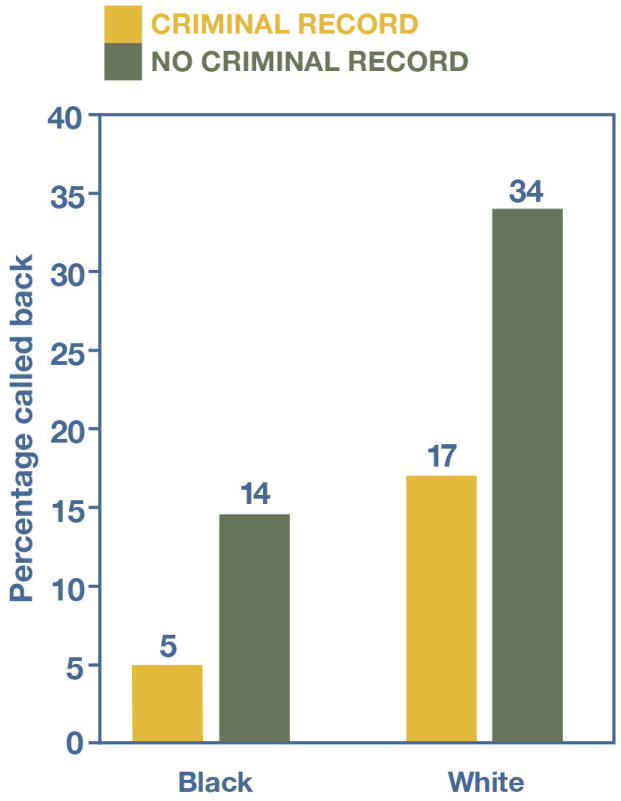


Figure 12.9. The Effect of a Criminal Record on Black and White Job Applicants: An Audit Study by Pager (2003). Devah Pager (2003) conducted an audit study regarding hiring practices for people with a criminal record. She hired four college students, two black and two white, to apply for a wide variety of entry-level local jobs. She found that whites with a criminal record were more likely to receive a callback than blacks—not only those with a criminal record but also those without one. *Figure adapted from a chart in Pager (2003)*

Pager and her colleagues reinforced and expanded on these findings in subsequent studies. For example, in a comparable audit study that she, Bart Bonikowski, and Bruce Western (2009) conducted in New York City, they hired white, African American, and Latino men to pretend to be job applicants. As in Pager's previous study, the testers were matched on a variety of characteristics and trained in the audit protocol. Then the researchers sent the men out to apply for entry-level jobs—this time, the same jobs—with equivalent résumés. Again, they found that white applicants received far more callbacks, and both African American and Latino applicants without criminal records did no better than white applicants just released from prison. A particularly interesting aspect of this later study was its qualitative component: the multiracial group of testers wrote down field notes describing their encounters with the same employers, and that ethnographic data gave the researchers some insight into how employers were making their decisions about hiring in ways that disadvantaged nonwhite applicants. For instance, the nonwhite testers would report being told that a position was already filled, while the white testers would be offered the same job later on. Nonwhite applicants would be asked to fill lower-status positions from the ones they originally applied for (bussers rather than servers, for instance), while white applicants were at times upgraded on the spot (asked to take jobs, say, as managers). This study's mixed-methods approach exemplifies how experiments can deftly incorporate qualitative data like the observations of an audit study's testers, allowing researchers not only to infer causality but also to illuminate the causal mechanisms that might bring about a cause-effect relationship.

Pager's research was cut short by her death at 46 from pancreatic cancer, but the remarkable findings from her audit studies have influenced the ways we think about race, incarceration, and economic opportunity (Seelye 2018). The huge effect that race independently had on employment outcomes across these studies was hard for even skeptical writers to refute, given the robust experimental design that Pager and her collaborators utilized, which accounted for every relevant alternative explanation. And the persistent finding that a felony conviction for drug possession could dramatically curtail employment opportunities helped generate public pressure on behalf of the so-called ban-the-box movement, which seeks to convince employers to stop asking about felony records on job applications. Large corporations like Walmart and Home Depot have now removed this question from their applications, in part because of Pager's important work.

The Trade-offs of Conducting Field Experiments versus Laboratory Experiments



A large body of audit studies has found that employers discriminate against people of color who apply for jobs. A 2017 meta-analysis of this research found that measured levels of discrimination against African American jobseekers have not changed since 1989, while discrimination against Latinos has declined somewhat (Quillian et al. 2017). *RDNE Stock Project, via Pexels*

Now that we've seen the many types of field experiments that social scientists can profitably conduct, let's talk a bit more about how laboratory and field experiments differ from one another—and how they might complement one another. Recall the "motherhood penalty" study (Correll et al. 2007) that we described in the last section. However striking the findings of the lab experiment were, critics might say that getting undergraduate students to "play" at making hiring decisions doesn't give us a plausible sense of what actually goes on when actual hiring managers look over actual résumés. Among other things, the study participants didn't have work experience in the relevant industry, which might mean stereotypical thinking figured more into their decisions than it would for actual hiring managers.

Perhaps anticipating these criticisms, the study's authors actually ran a field experiment in addition to their lab experiment. In response to newspaper job listings in a Northeastern U.S. city, the researchers submitted two sets of fake résumés and cover letters on behalf of equally qualified job applicants, based on the templates already used for the

fictitious marketing director position. One of the candidates was randomly assigned to be flagged as a parent in the job materials, and the two candidates for each position were of the same apparent gender (as signaled in their names). As in the lab experiment, the cues about the candidates' parental status were subtle—for instance, the résumé for the parent indicated they were an officer in an elementary school PTA, while the résumé for the nonparent said they were an officer in a college alumni association. The researchers then waited to see if these fake applications received any invitations for a phone or in-person interview. Again, the researchers found that mothers were discriminated against—with nonmother applications receiving significantly more callbacks than mother applications. The field experiment, however, did not uncover comparable evidence of a "fatherhood bonus" to match the lab experiment's findings.

As we've pointed out many times, field experiments generally do a better job with external validity than lab experiments do. Yet they do so at the price of experimenter control—which, in turn, may make it impossible to use randomization, control groups, testing, and the other features of classical experiments that, as we described earlier, help maintain a study's internal validity. We can see some of these trade-offs at work in the separate field and lab experiments that the "motherhood penalty" researchers conducted, which they explicitly did to address each experiment's weaknesses and thereby provide a more convincing case that their research findings were valid. The field experiment showed that employers discriminated against mothers—the same general finding that the lab experiment uncovered, but one that could be more convincingly generalized, given that the field experiment gathered its data based on interactions with real employers deciding on real jobs. The lab experiment had lower external validity by its very nature, but the researchers' control over their research setting meant they could go much further in parsing out the reasons that, say, mother candidates fared worse than nonmother candidates in job evaluations. They could specifically and precisely compare how mothers and nonmothers compared on various measures of competence and commitment, rather than having to rely on the flat yes/no response provided by a job callback in the field experiment. In other words, the lab experiment was better able to get into the heads of the people making hiring decisions and understand the reasoning behind their decisions, whereas the field experiment, at least in this case, was limited to learning about the decisions alone.

Key Takeaways

1. Field experiments employ features of classical experimental design, though they often lack much or any control over the real-world settings where they are conducted.
2. Audit studies typically are used to test discrimination in the labor market, the real estate market, and other settings. The researchers will hire and train testers who are matched on all characteristics except one—for instance, gender or race—so that the study can determine whether the testers are treated differently based on that sole characteristic.

12.5. Quasi-experiments: Approximating True Experiments with Observational Data



Instead of randomly assigning participants to control and treatment groups, some studies rely on matching. One matching approach—*individual matching*—involves pairing participants with similar personal attributes, but then having just one participant receive the treatment. *cottonbro studio, via Pexels*

Learning Objectives

1. Define quasi-experiments and describe the different types of studies employing this design.
2. Describe the characteristics of natural experiments and explain how they are different from field experiments.

When true experiments are not possible, researchers often use *quasi-experimental designs*.[1] These studies are similar to true experiments, but they lack random assignment to experimental and control groups. As a result, they are vulnerable to many of the threats to internal validity we talked about—selection bias above all.

As we discussed earlier, social scientists who pursue field experiments often make do without random assignment or other features of a classical experimental design because the real-world setting doesn't afford them that degree of control. They may tolerate a quasi-experimental approach because they explicitly want to generate findings with higher external validity. Beyond that specific consideration, however, quasi-experimental designs often make a lot of sense for researchers.

For one thing, controlled experiments are demanding. Because participants are subjected to an experimental manipulation of some kind, experiments (whether conducted in a laboratory or in the field) require more of participants and are more intrusive in their lives than other research designs. As we discussed in **Chapter 8: Ethics**, the rules and guidelines that govern research today largely emerged in response to wildly unethical and dangerous experiments, including social scientific experiments. Sociologists who employ this methodological approach must be especially careful about following ethical principles to prevent abuses. These considerations figure not only into the informed consent process for recruiting participants, but also in the randomization process of assigning them to experimental and control groups—for instance, offering benefits based on a lottery, as with the MTO study, can be problematic depending on the type of assistance being offered. Many times, too, the decision about whether to use a classical experimental design is not up to us. We may lack the resources of time or money needed to go that route, or we may not have the institutional support we need to structure our experiment in the most ideal way.

In this section, we'll talk about various quasi-experimental approaches that researchers may pursue for ethical or pragmatic reasons. When we can't randomly assign subjects into control and experimental groups, the next best thing

1. At times, authors will use the terms "quasi-experiment" and "natural experiment" interchangeably. Others refer to quasi-experiments as a category *distinct* from natural experiments, rather than as an overarching category that includes interventional quasi-experiments and natural quasi-experiments; in this view, "quasi-experiments" refers only to interventional experiments in which there is no random assignment, but the researchers decide on the assignment of participants to the control and experimental groups (rather than having outside events "naturally" make those assignments). We think the typology we've described here is a bit more intuitive and easier to understand, but you should be aware of this academic debate.

for us to do is to approximate that random assignment. There are two popular strategies for doing this. First, we can assign treatment and control groups ourselves, using a wide range of techniques like the ones we will discuss next—from comparison groups and matching, to time-series and regression discontinuity designs. These quasi-experiments are much the same as all the other experiments we've been talking about so far, in that they are **interventional experiments**—that is, the researchers intervene in the experimental setting (here, to create the control and treatment groups). This approach can be contrasted with **natural experiments**. In this second kind of quasi-experiment, we creatively exploit events in the real world that serendipitously assign our participants to treatment and control groups for us. We describe this category of quasi-experiments in its own section. Note that some techniques used in interventional quasi-experiments, such as comparison groups and regression discontinuity, can be used in natural experiments as well. The key difference between these two types of quasi-experiments, again, is *who* assigned the two groups being compared—the researchers or "nature"?

A **nonequivalent control group design** is the most commonly used quasi-experimental design. As shown in **Figure 12.10**, this design amounts to a classical experiment without random assignment. Note that researchers may refer to their control groups as **comparison groups** when they use this design. The lack of randomization means they cannot be as confident that the two groups are as comparable as control and treatment groups are meant to be.

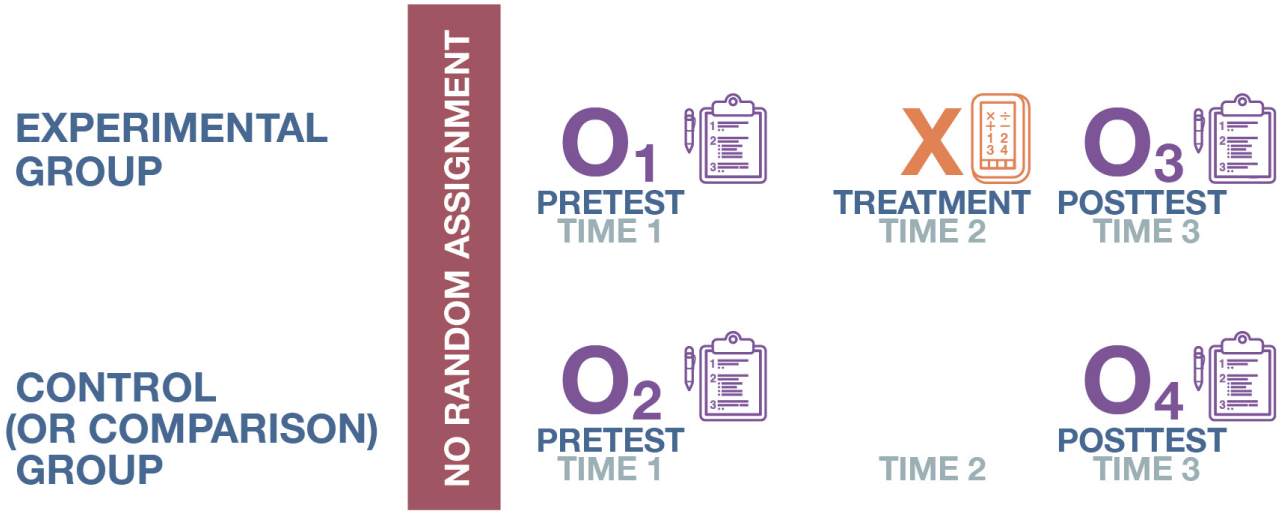


Figure 12.10. Nonequivalent Control Group Design. In this common quasi-experimental design, a study lacks random assignment to control and experimental groups. The selection threat to internal validity becomes more worrisome in this arrangement, but practical considerations may force researchers to treat two groups as roughly equivalent in this fashion.

Researchers sometimes choose this option because similar groups already exist that can serve as control and experimental groups without the need to go through the trouble of randomization. By doing so, they create a major vulnerability (i.e., selection threat) in their study, but practical considerations may overrule that fear. Indeed, when we were considering the different experimental designs we could use to evaluate the AddUpDog math app earlier, we talked rather blithely about how we would randomly select fourth-graders for our experimental and control groups. If we actually tried to conduct this experiment at an elementary school, however, administrators might very well reject this research design. They might tell us, for example, to provide the app to all the students in a particular classroom and to compare their test scores with those of students in another classroom.

Although school administrators might see this as a much more practical strategy to implement, it ruins our neatly randomized design. For example, perhaps the students in the classroom we selected as our treatment group are more

enthusiastic about extracurricular enrichment than the students in the classroom we selected as our control group. They may score equivalently on our math pretest, but their greater willingness to use the app may make it more effective in their hands, exaggerating our estimate of the app's ability to improve math skills for the general fourth-grade population. This selection bias is always a possibility when we can't use randomization to ensure equivalent groups. Unless we can adequately account for all the relevant characteristics that might influence the dependent variable and also differ across our two groups, selection will remain a potential explanation for our results.

A similar quasi-experimental approach replaces random assignment with a **matching procedure**. Two kinds of matching are commonly used. **Individual matching** involves pairing participants with similar personal attributes. In the math app example, we could pair two students who exhibit a high appreciation for math. The matched pair is then randomly split, with one participant going to the experimental group and the other to the control group. We would continue with this process until we have enough participants in our two groups. In **aggregate matching**, researchers effectively adopt a nonequivalent control group design, identifying a comparison group that is similar on important variables to the experimental group. Multiple pairs of treatment and comparison groups can be created in this fashion. For instance, in our AddUpDog evaluation, let's say we give various fourth-grade classrooms a survey measuring their appreciation for math. Then we match two classrooms with high levels of appreciation, two more with moderate levels, and two more with low levels. Finally, we randomly assign one of the two classrooms at each level to the experimental group and the other to the control group.

To increase internal validity, researchers using matching procedures have to think carefully about which variables are important in their study, particularly when it comes to demographic variables or attributes that might affect their dependent variable. Those are the variables they will want to use for matching purposes. As you might have guessed, the downside here is that we need to know beforehand which variables are important. We are only able to match on characteristics that we know about. It is possible we might miss something relevant—such as the enthusiasm gap we noted earlier across a school's fourth-grade classrooms. We might diligently match, only to find out later that we should have chosen other characteristics to ensure that we had truly comparable groups. Randomization makes that task far easier, and choosing matching in its stead means we will always have to be on the lookout for possibly decisive differences in our groups. To their credit, quantitative researchers have developed sophisticated procedures to match cases based on multiple variables, creating **propensity scores** that quantify how similar cases are across those variables and automatically matching cases according to each one's propensity score. At the end of the day, though, the essential problem remains that we need to make sure we have accounted for all the variables that might matter—all the possible confounders that might explain our results.

We should note that longitudinal **panel studies** are sometimes described as quasi-experiments because they basically involve matching, too. When we have data over time on the same individual, we can see whether their outcomes change because of the introduction of the independent variable—say, if their income increases after attending a training program. The person before the treatment is essentially our control group, and the person post-treatment (after going through the training) is our experimental group. The longitudinal design rules out many alternative explanations for the observed changes in the dependent variable because we're looking at the exact same person. This is why sociologists are more confident about making causal arguments when they have longitudinal rather than cross-sectional data.

A more advanced type of quasi-experiment is a **regression discontinuity design**. In this approach, researchers choose their treatment and control groups from cases clustered around a "cutoff" point (a discontinuity) in a particular variable that determined whether the treatment was given. To give you a sense of what this technique entails, consider a study conducted by the World Bank, an intergovernmental organization that provides financing to lower-income countries pursuing development projects. World Bank researchers Deon Filmer and Norbert Schady (2009) wanted to evaluate whether a program that provided scholarships to at-risk children in Cambodia actually improved their school enrollment and attendance. They had data on the children who received the scholarship and those who didn't, but they obviously couldn't just compare school enrollment and attendance across the two groups, which differed on a variety of

characteristics that could themselves explain any divergences in those outcomes. Furthermore, the program had not set up a lottery for distributing its scholarship (like the MTO experiment had done), which would have made measuring the impact of the scholarship more straightforward. That said, the sample of students available to them was so large that the researchers could pursue a regression discontinuity design. The program had given the scholarship to at-risk children, identifying them based on a "dropout risk score" they calculated using data on household characteristics and student test scores. What if researchers looked *just* at those students who were on either side of the cutoff for receiving the scholarship—those who scored right above and below the risk score threshold for eligibility? These students would be quite similar to each other, with a minor difference in their dropout risk score determining whether or not they received the scholarship. Thanks to their large sample, researchers were able to analyze outcomes just for those children whose dropout risk scores fell within a narrow 10-point band around the cutoff point. Within this fairly comparable group of students, they found that those children who received the scholarship had substantially higher school enrollment and attendance. You can see this striking discontinuity in **Figure 12.11**, where the vertical line shows a steep drop in educational attainment after students are no longer eligible for the scholarship.

You can appreciate why regression discontinuity designs require such large samples: researchers must focus their analysis on the cases clustered around the cutoff point while dropping many others. This design also requires a cutoff point that does not coincide with other important cutoffs (say, eligibility for other government programs) so that the participants on either side of that line are truly similar to one another except for the intervention of the stimulus.

With all the checks and balances provided by random assignment, the results of true experiments are much easier to defend than those of quasi-experiments—which detractors have dubbed "queasy experiments," given the warranted fears that researchers should have about their internal validity. That said, regression discontinuity designs in particular have been shown to obtain similar results to studies using classic experimental designs. And for all their vulnerabilities, quasi-experiments have generated a great deal of compelling new research over the past few decades, enabling social scientists to study research questions that are tough or impossible to examine in a randomized controlled experiment. Indeed, an especially vibrant area of recent research involves our second strategy—quasi-experiments in which the assignment of control and treatment groups is entirely out of the researchers' hands.

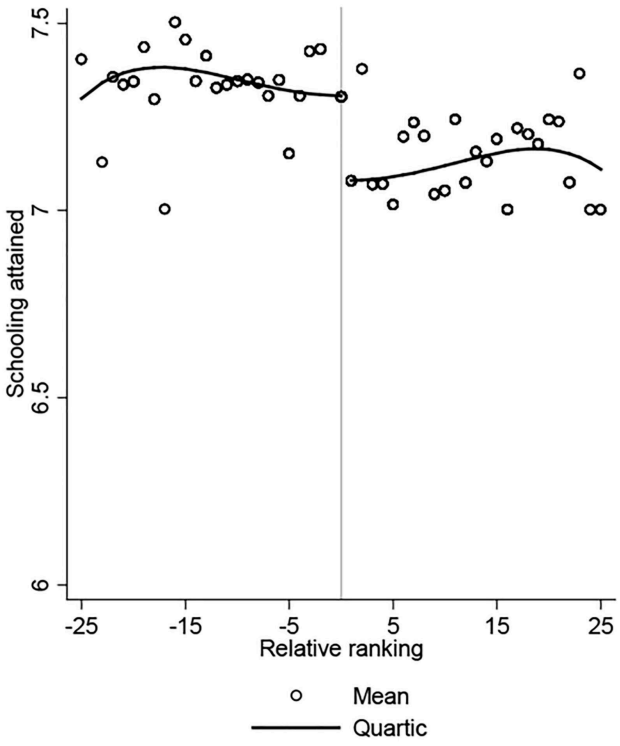


Figure 12.11. Educational Attainment of Cambodian Students by Dropout Risk Score: Regression Discontinuity Analysis by Filmer and Schady (2009). In a *regression discontinuity design*, researchers choose their treatment and control groups from cases clustered around a "cutoff" point (a discontinuity) in a particular variable that determined whether the treatment was given. This chart comes from a World Bank evaluation of a Cambodian scholarship program for at-risk children. Using a regression discontinuity analysis, researchers compared the (highly similar) students on either side of the cutoff for receiving the scholarship and found that those who received it had substantially higher school enrollment and attendance. *Deon Filmer and Norbert Schady, via World Bank*

Natural Experiments



Natural disasters can become the basis for natural experiments, as David Kirk (2009) showed in a study that examined whether returning to their former neighborhoods made parolees more likely to offend again. Because of the destruction wreaked by Hurricane Katrina in 2005, many Louisiana parolees could not go back to their previous homes and had to move out of their parishes. Using this random consequence of a natural disaster to test his hypothesis, Kirk found that relocating did lower rates of recidivism. *Franklin Peña Gutierrez, via Pexels*

Sometimes events in the real world happen to provide a researcher with novel ways to introduce features of experimental design into the analysis of observational data. In these particular cases, the researcher has no control over the event itself, which could be anything from a natural disaster, to a change in the weather, to the enactment of a new government policy. The important thing is that the occurrence of the event serves as a de facto "treatment," changing the independent variable of interest. The researcher designs their study around the existing characteristics of the situation.

This sort of study is called a *natural experiment*. In this context, "natural" does not mean the study has to involve a natural disaster (though a surprising number of social scientific studies do make use of hurricanes and the like for this purpose!). Rather, it indicates that the study's researchers did not have control over the experiment. They did not themselves manipulate the independent variable. They could not create (artificial) conditions to precisely test the stimulus or treatment as they saw fit. This lack of control distinguishes natural experiments from interventional experiments. In fact, it's probably more accurate to say that researchers "exploit" a natural experiment rather than

"conduct" it. Conditions in the real world are changing on their own, and the savvy researcher is taking advantage of a specific configuration of changes for their research purposes—sometimes long after the event transpired.

One of the most influential natural experiments in recent decades was a study by David Card and Alan Krueger (1994) of the effect of minimum-wage laws on employment. Before this study, economists had tightly held to the notion that any increase in the minimum wage—the lowest hourly wage that employers could pay their workers, by law—would generate greater unemployment. After all, economists theorized, employers would need to pay each worker more, and since they did not want to lower profits or raise prices—which would drive away customers—they would likely cut hours and lay off employees.

How could this theory be put to the test? Simply analyzing changes in employment after a state or locality decided to increase its minimum wage was fraught with problems, given all the possible alternative explanations we discussed earlier in the chapter. And for ethical and practical reasons, it was impossible to conduct a laboratory or field experiment that could persuasively assess this hypothesized causal relationship. Card and Krueger, however, decided to compare employment levels in two states—New Jersey and Pennsylvania—after New Jersey decided in 1992 to raise its minimum wage (from $4.25 to $5.05). In this natural experiment, New Jersey was essentially the treatment group, and Pennsylvania (which did not raise its minimum wage) was the control group.

Card and Krueger narrowed their focus to make the comparison more akin to a true experiment. First, they obtained employment levels (measured by hours worked on a full-time basis) for fast-food establishments. Fast-food restaurants typically pay their workers the minimum wage, so these businesses should be most sensitive to any changes in minimum-wage laws. Second, the researchers collected data from a more restricted geographic area. Had they looked at statewide employment levels for New Jersey and Pennsylvania, the comparison would have been apples-to-oranges: any two states are quite different in terms of their demographics and industrial makeup, which could be responsible for any observed differences in employment outcomes over time. To address this issue, Card and Krueger focused on fast-food establishments along the border of the two states, with the idea that the restaurants on either side of the border would be situated in similar communities whose main difference was which state policies affected them.

When Card and Kruger narrowed their analysis in these ways, they found that in the months before and after April 1, 1992—the day the minimum wage rose by 80 cents in New Jersey—fast-food employment rose in New Jersey but fell in Pennsylvania. These trends ran contrary to what prevailing economic theory predicted—that New Jersey should see a hit to employment after hiking its minimum wage.

As we have discussed, however, a true experimental design assumes that the experimental and control groups are equivalent across all relevant characteristics except for the experimental group receiving the treatment. Although comparing fast-food establishments across the border helped make the two groups comparable, there was still a substantial difference in employment levels between the two areas before the law change. New Jersey's employment in the sector was actually lower than Pennsylvania's during this earlier period. These different starting points—divergent pretest scores, essentially—needed to be accounted for. To do so, Card and Krueger employed a statistical technique called **differences-in-differences** (commonly used in studies exploiting natural experiments). As shown in **Figure 12.12**, this approach first calculates the change over time in the dependent variable for the treatment group (the treatment group's "difference" over time) and then calculates the change over time for the control group (the control group's "difference" over time). Then the two "differences" are compared ("differences in differences"). This technique accounts for the fact that the treatment and control groups did not start at the same place on the pretest. Had New Jersey not passed its minimum-wage law, its employment levels should have followed the downward path that Pennsylvania took, but starting from an even lower level. This counterfactual scenario, illustrated by the blue dashed line, dips down at the same angle as Pennsylvania's employment does, landing at a point lower than the point at which Pennsylvania arrived in the end (Point A). Therefore, a proper measure of average treatment effect—in this case, the impact of the minimum-wage increase on employment levels—would be the difference between that hypothetical scenario, the counterfactual (Point A), and what New Jersey actually experienced (Point B). When the researchers did that analysis, they found the

increase in employment apparently caused by the minimum-wage hike to be even higher than the New Jersey trend suggested on its own.

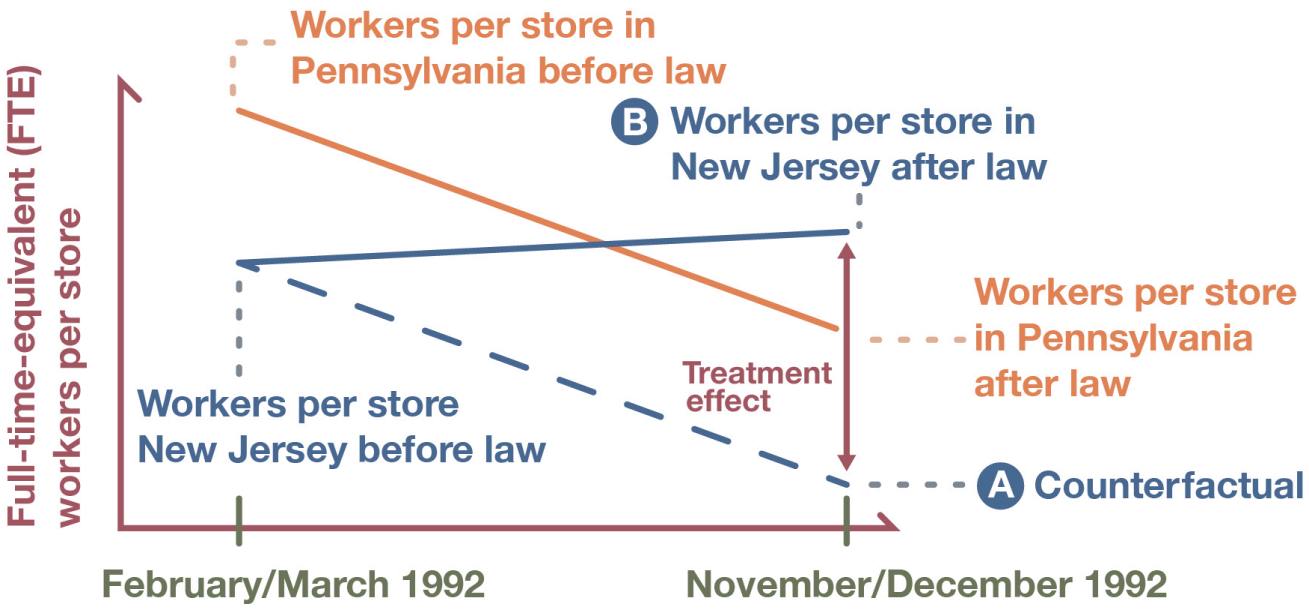


Figure 12.12. The Logic of the Differences-in-Differences Method in a Minimum-Wage Study by Card and Krueger (1994). The economists David Card and Alan Krueger (1994) overturned the conventional wisdom about minimum-wage increases through a quasi-experimental comparison they conducted of employment levels in two states—New Jersey and Pennsylvania—before and after New Jersey raised its minimum wage. Contrary to prevailing economic theory, they found that employment levels did not drop in New Jersey following its minimum-wage increase. In fact, after accounting for the higher level of employment in Pennsylvania before the wage hike (using a technique called *differences-in-differences*), Card and Krueger concluded that the minimum-wage law led to a substantial increase in employment in New Jersey. In this analysis, note that full-time-equivalent (FTE) employment is measured as the number of full-time workers (including managers) plus half the number of part-time workers; however, the diagram is meant to illustrate the overall relationships and is not drawn to scale.

Sometimes the "natural" intervention that researchers exploit in a natural experiment is exceedingly random and would seem to have little bearing on the dependent variable of interest—which is often the point, given the desire to avoid a selection effect. Sociologist David Kirk (2009) used a natural experiment to study whether people released from prison were less likely to return if they moved away from their former neighborhoods. The theory was that relocating would help break a parolee's ties to friends and acquaintances involved with the sorts of criminal activity that had earlier landed the person in prison. Of course, a person's housing decisions aren't something that can be easily or ethically manipulated in a lab or field setting: we couldn't, for example, randomly choose some of the parolees to relocate and others to stay in their previous homes, and then track their recidivism (return to prison) rates. And if we wanted to just use observational data—say, comparing the recidivism rate of parolees who move after their release with those who do not—we would be concerned about the selection threat. Parolees who change their residence may be systematically different from those who do not—for instance, the movers may be more financially secure—and those differences may make a return to prison more or less likely. In short, an observational analysis would not be able to adjudicate between the two causal stories that (1) relocation actually made some people less likely to return to prison, or (2) the two groups were just different to begin with.



Patrick Meirick and Stephanie Schartel Dunn (2015) did not have to rely on a natural disaster for their natural experiment. Instead, they used Barack Obama's appearances in televised debates during the 2008 presidential campaign to see whether the presence of a high-profile, charismatic African American leader literally on the national stage would shift Americans' opinions about race. Tracking attitudes before and after the debates through a survey fielded multiple times over that period, the researchers found that watching Obama in the debates did improve people's attitudes toward African Americans, though the effects were weak. *Jose A. Torres Jr./U.S. Department of Defense, via Wikimedia Commons*

Kirk found an ingenious natural experiment he could use to test his hypothesis connecting relocation to lower recidivism. Hurricane Katrina, a natural disaster that devastated New Orleans and other coastal areas in 2005, had leveled hundreds of thousands of homes. Kirk decided to study the recidivism rates of Louisiana parolees who were released in the period immediately following the destructive hurricane. If a parolee's residence was severely damaged or destroyed, they had to move. Furthermore, the hurricane's impact on parolees was random and had no relation to characteristics affecting their likelihood of recidivism. Therefore, parolees who had lost their homes could serve as a plausible treatment group for the study, capturing the effect of relocation without as great of a risk of selection. As an additional check, Kirk statistically controlled for multiple variables that could influence recidivism. (We'll say more about how statistical controls can approximate control groups at the end of the chapter.) In line with his hypothesis, Kirk's analysis found that parolees were significantly less likely to return to prison if they moved out of the parish.

To exploit natural experiments, researchers often need longitudinal data spanning the time before and after the event or policy change that serves as their treatment. For instance, Patrick Meirick and Stephanie Schartel Dunn (2015) wanted to see if Americans' opinions about race changed as a result of Barack Obama's 2008 presidential campaign. They focused on a series of events—the televised debates between the Democratic nominee and his Republican challenger John

McCain—which, they argued, "gave tens of millions of Americans an extended look at a highly vivid, salient, cognitively available, counterstereotypical exemplar that may have changed the way they feel about African Americans" (Meirick and Schartel Dunn 2015:58). As luck would have it, a telephone-recruited internet panel of the American National Election Survey (ANES) included a battery of questions about racial attitudes that they could use, and the panel had been surveyed multiple times that year, including in the months before and after the debates. Furthermore, the November survey asked respondents whether they had watched the debates, allowing the researchers to identify a comparison group of people who did not see the debates.

As you might be thinking, the natural experiment that Meirick and Dunn exploited was far from perfect. Among other things, their treatment and control groups were nonequivalent—we might expect non-debate watchers to be different from debate watchers on a host of characteristics. (To make the two groups more comparable, the researchers used statistical controls for demographic and party identification variables.) Nevertheless, the fortuitous availability of the ANES data and timing of the debates allowed the researchers to make a tidy comparison of pre- and post-debate racial attitudes. They concluded that watching the presidential debates led people to feel more positive affect toward African Americans across a range of explicit and implicit measures. The relationships, however, were relatively weak.

As you can see across these examples, making use of a natural experimental design requires an ample amount of imaginative and versatile thinking. Researchers need to spot impactful (though possibly inconspicuous) developments in the world around them. They need to thoughtfully leverage those occurrences to identify treatment and control groups. And they often need to be exceedingly lucky, not just in terms of being around when the right disaster or policy change occurs, but also in terms of having access to data spanning the event in question. As daunting as this research endeavor can be, it allows researchers to investigate elusive connections between variables, extending experimental approaches to new domains while possibly avoiding the selection effects that routinely plague the analysis of observational data.

It's also worth emphasizing how natural experiments are on the other end of the spectrum from lab experiments when it comes to experimenter control. Outside events determine the structure of the experiment, not the researcher's own decisions. That means the researcher may find it difficult to comprehensively assess the effects (if any) of the treatment on outcomes—they may be stuck with whatever data happened to be collected. It also means that natural experiments are impossible to replicate: you'll have to wait for another disaster to strike or another law to get enacted. We're being a bit facetious here—clearly, one could design similar studies to test the same hypotheses—but replicating such a study is much harder than simply following a lab experiment's methodological recipe. It is therefore harder to be confident that the results of natural experiments are not the result of some fluke—however savvily that fluke was exploited for research purposes.

Deeper Dive: Time-Series Design

Earlier in the chapter, we noted how the absence of a control group posed major problems for an experiment's internal validity, including the possibility that concurrent events could be responsible for observed changes (history threat) or the passage of time could lead to natural changes in outcomes (maturation threat). Although creating a control group alongside the experimental group remains the best way to deal with these threats, to some extent, they can be lessened by conducting several pretests and posttests over a longer period of time. We call this quasi-experimental approach—which has neither randomization nor a control group—a **time-series design**.



Drawing from birth-rate data immediately before and after the 1995 Oklahoma City bombing, sociologists used a *time-series design* to test theories that experiencing a disaster prompts people to have more children. *Preston Chasteen, via Wikimedia Commons*

Let's say that for whatever reason we cannot recruit a control group for our evaluation of the AddUpDog app—no child wants to miss the chance to touch (and tap) such technological greatness. We decide to use a time-series design as an alternative. As shown in **Figure 12.13**, we might decide to do four pretests before the students start using the app and then four posttests after they complete their 10 hours. How does this times-series design help us conduct a rigorous analysis of whether using the app improves math skills?



Figure 12.13. One-Group Time-Series Design. In a time-series design, there is no control group, but researchers make multiple observations both before and after the stimulus has occurred and see whether outcomes differ substantially across the two time periods.

In comparison to just using one pretest and one posttest, this experimental design gives us a good sense of how stable the students' scores were both before and after they used the app. Remember that a major threat to internal validity when we don't have a control group is that other factors outside the experiment might account for abnormally high or low values on our experimental group's pretest or posttest, which we could mistake for the effect of the treatment. With multiple pretests and posttests, these extraneous factors should not be as much of a concern, washing out across our many tests. Furthermore, this time-series design allows us to see whether the students' posttest scores increased right after they used the app and whether they stayed high. If these two conditions are true, then we have some evidence that the app truly caused improvements in their math skills: the timing of the spike and its lasting effect both suggest a causal relationship. **Figure 12.14** depicts what the trend lines would look like when the stimulus truly does have a causal effect.

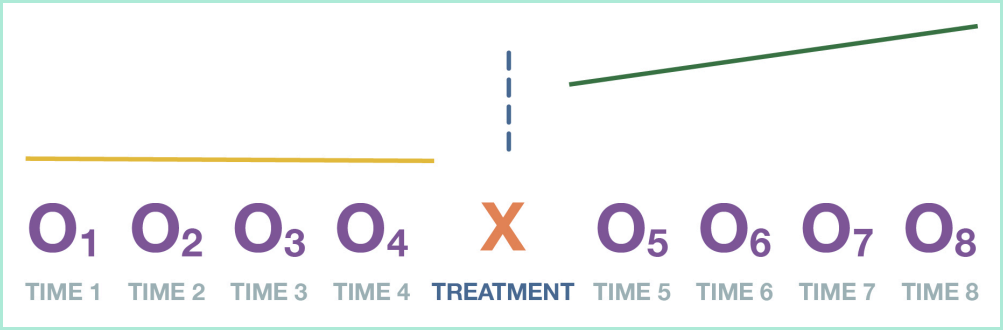


Figure 12.14. Time-Series Outcome When the Treatment Has an Effect. In a time-series design, we would expect to see the trend line change dramatically—as it does in this diagram—if it truly does have an effect on the outcome of interest.

In short, what we'd like to see if our math app improves students' test scores is some kind of clear discontinuity between the trend in scores before and after the stimulus. That pattern would suggest that the math app really did make a difference.

A time-series design has multiple observations of the dependent variable both before and after the introduction of the stimulus, but the exact number is up to the researcher—the more pretests and posttests, the clearer the trend lines become. With this design, the researcher can better understand the true value of the dependent variable before any experimental intervention. Likewise, the multiple observations afterward allow the researcher to compare the before and after trends in the dependent variable.

A study by sociologists Joseph Rodgers, Craig St. John, and Ronnie Coleman (2005) used a time-series design to test a set of social psychological theories that argued people had more children in response to natural disasters and disruptive political and cultural events. For their creative quasi-experiment, the researchers drew from birth-rate data immediately before and after the 1995 bombing of the Murrah Federal Building in Oklahoma City, a terrorist attack that killed 168 people, including 19 children. Their hypothesis was that local birth rates should have increased immediately afterward, consistent with the theoretical view that disasters—by drawing attention to people's mortality—prompted them to hew to more traditional values and practices like having families and raising children.

When the researchers used advanced statistical techniques to graph their time-series data (shown in **Figure 12.15**), they saw a great deal of fluctuation in the trend line, as is typical for this kind of demographic data. Notably, while the fertility rate in Oklahoma County—where Oklahoma City is located—varied before the

bombing (1990 to early 1996), it seemed to vary around a relatively flat line.[2] By contrast, in the period after the bombing (early 1996 to 1999), the fertility rate sloped upward. This discontinuity in the trend line suggested that the bombing did spur an increase in births, in line with social psychological theories.

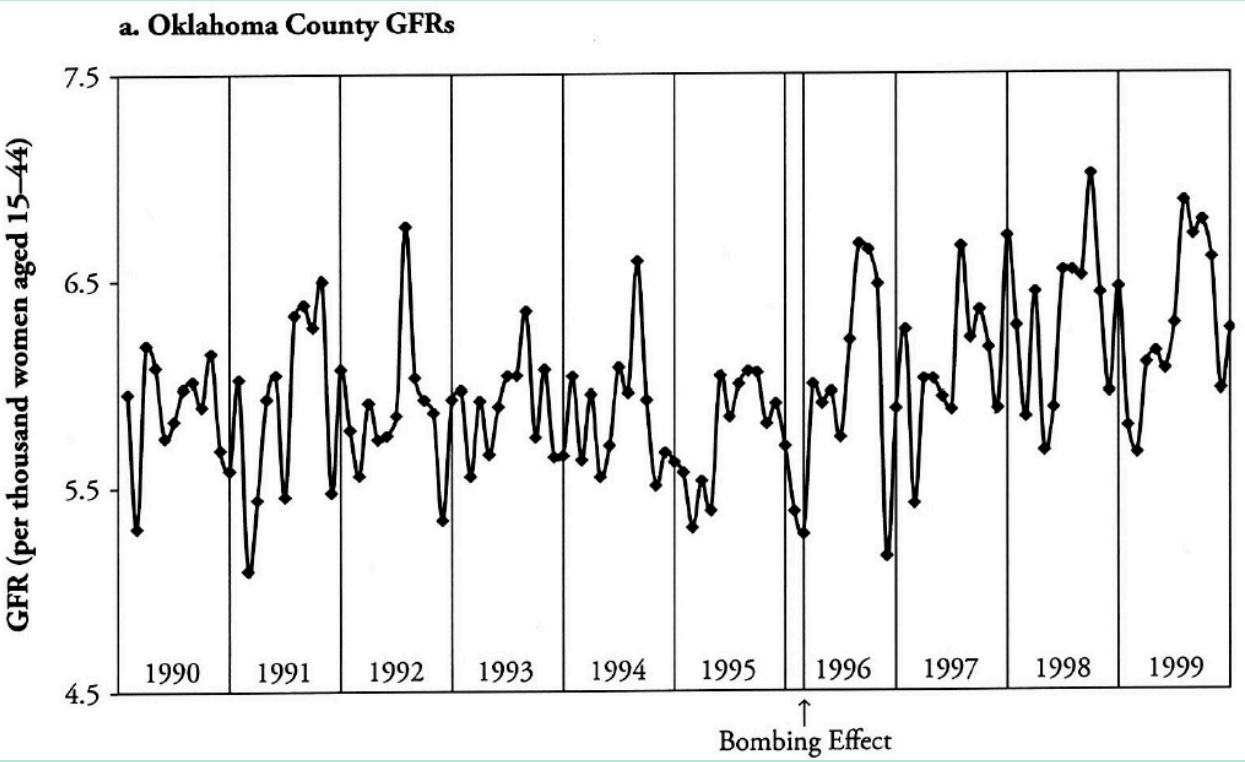


Figure 12.15. Oklahoma County General Fertility Rate over Time: A Time-Series Experimental Design (Rodgers et al. 2005). The local fertility rate trended upward after the 1995 Oklahoma City bombing, in line with theories that disasters cause births to surge. *Figure from Rodgers et al. (2005) with annotation*

A time-series design can be useful when a more rigorous experimental approach cannot be taken, but it remains quite vulnerable to threats to internal validity. Even if the slopes of the trend lines we observe seem to confirm our hypothesis, we have to be wary about alternative explanations that have not been ruled out through the use of a control group and randomized assignment. For instance, while the post-bombing increase in Oklahoma County's birth rate is striking, it may be the case that another event happening around the same time as the bombing was responsible for the increase (history threat). Or perhaps the fertility rate happened to be extraordinarily low in the period before the bombing and regressed to its mean afterward (regression threat).

2. We have drawn two separate lines on this graph to represent an approximation of an average of the points. This is a rough visual method. If we had the original dataset, we could have our data analysis program calculate a least squares regression line, which would provide a more accurate depiction.

Having a control group would strengthen any finding that the Oklahoma City bombing increased the fertility rate. But that was not possible in this retrospective study—and obviously, exposing people to trauma to see the effect on their decisions to have children is not ethical or feasible. It is possible, however, to add one or more comparison groups to a time-series design, creating what is called a **multiple time-series design**. The approach is depicted in **Figure 12.16**.



Figure 12.16. Multiple Time-Series Design with Two Comparison Groups. To further test the theory that disasters increase fertility, Rodgers and his collaborators (2005) employed a multiple time-series design. If the theory was true, they hypothesized, then Oklahoma County—the county where the bombing occurred—should have experienced a greater surge in births than other counties in the state did. Those other counties served as comparison groups in their analysis, which found that the spike in fertility occurred in and around the Oklahoma City area but not elsewhere—thereby providing more evidence in support of the theory.

Rodgers and his collaborators (2005) used a multiple time-series design to strengthen the internal validity of their study. How did they go about identifying comparison groups? They reasoned that if the theory connecting disasters to higher fertility were true, Oklahoma County, the county where the bombing occurred, should have experienced a higher increase in fertility afterward than other counties in the state did. Least able to tune out the horrifying consequences of the terrorist attack, Oklahoma City residents would have been the group most

likely to follow through on the theory's predictions of having more children. With this hypothesis in mind, the researchers used the other counties as comparison groups and saw how their fertility trend lines compared with those of Oklahoma County. They found that a baby boom appeared largely in and around the Oklahoma City area and persisted throughout the county, but the same dynamics did not hold elsewhere. Through the use of comparison groups, the scholars studying the Oklahoma City bombing were better able to support their causal argument that the disaster truly did lead to an increase in births.

Key Takeaways

1. The defining characteristic of a quasi-experimental study is its inability to randomize experimental and control groups.
2. Different types of quasi-experimental designs include nonequivalent control group design, panel study, regression discontinuity design, and time-series design.
3. A natural experiment is one in which the researcher does not manipulate the independent variable. Instead, a naturally occurring event essentially creates treatment and control groups, allowing researchers to compare outcomes across them.

Exercise

We have distinguished between laboratory and field experiments, true experiments and quasi-experiments, and interventional experiments and natural experiments. Refer to the examples of studies we have mentioned throughout this chapter. How would you categorize each one according to these three sets of categories?

12.6. Experiments: Our Best Way of Inferring Causality, but Far from Foolproof



An influential lab experiment that plied participants with chocolate chip cookies and radishes developed the psychological theory of "ego depletion," which holds that forcing people to resist temptation or make hard decisions exhausts the brain and makes it harder to exercise willpower or choose wisely later in the day (Baumeister et al. 1998). While hundreds of psychological studies built on this theory over the next two decades, two massive studies drawing from the cooperation of dozens of labs across the world stunned the research community by finding no evidence of an ego-depletion effect. *cottonbro studio*, *via* *Pexels*

Learning Objectives

1. Discuss the "replication crisis" and its implications for social scientific research.
2. Understand how publication bias tends to discourage replication efforts in social science.

We've gone over the many ways that experimental designs can fall short of the classical experimental ideal. We've also talked about the inveterate problems of external validity that plague experiments, even if field experiments can offer a middle ground of sorts in producing generalizable results. It is worth emphasizing again that even when we are able to do an experiment with randomized control groups, pretests and posttests, and a precisely manipulated treatment, we can still run into problems with internal validity.

In fact, social scientists are coming to recognize how even well-executed experiments can fall short, particular given the growing evidence that many experimental results published in peer-reviewed journals cannot be replicated. As part of an initiative called the Reproducibility Project, researchers from around the world replicated 100 studies from three top psychology journals (Resnick 2018). More than half of these empirical studies—many of which had used the most rigorous experimental designs—failed to pass muster: the project's researchers could not reproduce statistically significant results in line with the original studies' findings. For whatever reasons, the accountability systems built into peer-reviewed science had not succeeded in weeding out these studies earlier—perhaps because those studies had not been conducted properly, but also perhaps because some of the more striking findings had been dumb luck, and no one had bothered to ensure that they held up.

One of the most startling casualties of recent replication efforts has been the psychological theory of "ego depletion." A research team led by Roy Baumeister and Dianne Tice seated participants before a table with a stack of just-baked chocolate chip cookies and a bowl of plain red and white radishes (Baumeister et al. 1998). The researchers found that an experimental group forced to resist the cookies and eat the radishes instead gave up on an impossible-to-solve geometric puzzle much faster than a control group given free range to chow down on the sugary treats. This study led to a vibrant line of research around theories of "ego depletion" and "decision fatigue," which view willpower as a muscle, with repeated instances of resisting temptations and making tough decisions sapping the brain's limited capacity for self-control and decision-making. Hundreds of peer-reviewed studies built on this foundation—measuring the consequences of mental exhaustion for everything from altruistic behaviors to athletic performance—and the ideas trickled into public consciousness, with even Barack Obama reportedly avoiding the morning ritual of personally picking his suits to avoid decision fatigue (Resnick 2016). Yet a large-scale, systematic replication involving 2,141 participants in multiple continents debunked the theory of ego depletion when measured using the exact same test across those sites (Hagger et al. 2016). Another multilab replication involving 36 sites and 3,531 participants also failed to produce evidence of any ego-depletion effect (Vohs et al. 2021). "I'm in a dark place," one of the first replication's contributors, a psychologist who specializes in ego depletion, wrote on his blog (Inzlicht 2016). "I feel like the ground is moving from underneath me and I no longer know what is real and what is not."

In response to this so-called replication crisis, some scientists have raised concerns that today's conventions regarding reviewing scholarly work and publishing it may be contributing to a larger problem of reproducibility, whereby the checks and balances built into the scientific system—peer review and replication—aren't working as well as they should be. In a 2016 poll by the prestigious scientific journal *Nature* of 1,500 scientists from various fields, 70 percent said they had failed to replicate at least one other scientist's work—and 50 percent admitted they had failed to reproduce their own work. Only a minority of these scientists, however, had ever sought to publish a replication in a journal. In general, academic journals seek to publish new, trailblazing work rather than paint-by-numbers replications, so academics have few incentives to rehash what is seen as settled science. Indeed, in their feedback, several respondents in the *Nature* poll brought up instances when they had succeeded in publishing failed replications in the face of pushback from journal editors and reviewers, who had asked the authors to downplay how their research contradicted the original studies.

Note that **publication bias**—the tendency of journals to publish work drawing provocative connections between phenomena while passing on papers that merely test earlier findings—is an issue that plagues scientific work of all kinds, not just experimental studies. In fact, one of the key advantages of experiments as a research method is that they are relatively easy to replicate. Individual experiments, particularly laboratory experiments limited in scope, often require relatively little investment in time and other resources. That's not the case for immersive ethnographic studies

or wide-ranging surveys. And any one study—even a well-designed one—can produce results that happen to be a fluke, so replication is essential to keep science honest. (By the way, this is another reason that you shouldn't pay too much attention to every new headline about a study showing major health benefits or dangers of eating one food or another—wait to see those findings replicated before you overhaul your way of life.) It's likely that a Reproducibility Project focused on vetting sociological studies would find even more grim news, given that sociology does not rely as heavily on experimental designs as psychology does, and even less of this sort of exacting replication occurs in our field.

We should be deeply concerned about the possibility that much social scientific research cannot be reproduced, given that replicability is one of the foundations of what makes a science a science, as we discussed in **Chapter 1: Introduction**. That said, it does no good to simply throw up our hands and reject science, the best method we have to observe and vet the truths of our reality. And its flaws notwithstanding, a robust experimental design remains the best scientific tool for inferring causality. Indeed, in the next two chapters, we'll discuss the ways that sociologists use statistical methods to essentially approximate the randomized treatment and control groups in experiments. As we noted in this chapter, quasi-experiments are themselves approximations of the key internal validity-preserving features of randomized controlled trials. They try their best to infer causality, but they often fall short of the classical experimental ideal.

We'll conclude this chapter by breaking down just how much more an experiment like the "motherhood penalty" study (Correll et al. 2007) can tell us about what is truly driving the outcomes we observe than just relying on observational data from surveys. As you probably know, women get paid less than men. A 2020 Pew Research Center report put this ratio at 84 cents for women for every dollar a man makes (calculated based on the median hourly earnings of both full- and part-time workers). Clearly, the observational data Pew is using here—and similar surveys of payroll data—show a stark gap in terms of how much men and women are paid, even accounting for the number of hours they work. But what factor is decisive in creating this gap? Is this 84 percent figure a good measure of how much employers discriminate against women? How much does it reflect differences in education and skill, or years of experience, or choices to pursue occupations that pay low wages or prioritize family over career? In short, what are we really measuring here? A similar issue complicates our analysis of the pay gap between mothers and nonmothers. Mothers clearly make less money over their lifetime, according to payroll data, but we can tell more than one causal story about why that is so. For one thing, is it that mothers truly are less committed to their jobs than nonmothers—hurting their performance and resulting in fewer promotions and less pay—or is it because employers *perceive* mothers to be less committed, and discriminate against them based on stereotypes even when they perform just as well?

If we just look at the patterns in cross-sectional observational data like this, we can't say all that much. We can't manipulate the independent variable. Our data is from one point in time. To deal with these issues, as we will explain in the next two chapters, social scientists will often use statistical controls when analyzing observational data. They use math and statistical theory to screen out explanations other than the independent variable of interest, essentially creating experimental and control groups within their samples. The problem is—as we alluded to in our discussion of matching, which follows a similar logic—a researcher following this approach must control for all confounders, and measure those confounders well. If they fail to do so, they can easily fall into the trap of selection bias and other threats to internal validity.

In the case of the gender pay gap and the motherhood penalty, it was exceedingly hard for social scientists to measure each of the possible factors at play and convincingly argue that they had isolated the effect of employer discrimination. That is why the study by Correll and her collaborators was so ingenious: they were able to create conditions in their lab and field experimental settings that ensured the only factor being tested was employee discrimination, and not hard-to-measure differences in the skill levels of the job candidates. They used their control over the study conditions to keep everything constant ("all other things being equal") except for the independent variable they were manipulating. The result was potent evidence that, yes, employee discrimination against mothers is real.

Although not as ideal from an internal validity standpoint, quantitative analysis of observational data from surveys and

other sources can be illuminating in its own right, as we discuss in the next two chapters. What is lost in internal validity may be gained in external validity, since surveys typically elicit people's opinions about real issues and situations rather than contrived scenarios. Rigorous statistical techniques can compensate for the lack of an experimental design. And ethical and pragmatic considerations often make true experiments hard to conduct, meaning that we sometimes have to make do with less-than-ideal methodological approaches. Indeed, some of the most interesting research of recent decades has been natural experiments and quasi-experiments that creatively examine phenomena that are otherwise difficult to study in a traditional experimental fashion—the idea here being, as with field experiments, that we might trade a degree of internal validity for feasibility or generalizability.

Personally, we believe that each of these methodological approaches has its merits. While controlled experiments remain the gold standard, we social scientists should use all the tools at our disposal. By triangulating with a variety of methods, we as a scientific community get steadily closer to being able to say with confidence that changes in one variable truly do, or do not, cause changes in another.

Key Takeaways

1. Researchers have failed to replicate findings of many influential psychological experiments, raising questions about social science's ability to infer causality.
2. Publication bias favors the publication of work drawing provocative connections between phenomena over papers that test earlier findings through replication, which may be contributing to the so-called replication crisis.

12.7. References

Baumeister, Roy E., Ellen Bratslavsky, Mark Muraven, and Dianne M. Tice. 1998. "Ego Depletion: Is the Active Self a Limited Resource?" *Journal of Personality and Social Psychology* 74(5):1252–65. doi: 10.1037/0022-3514.74.5.1252.

Becher, Emily H., Jenifer K. Mcguire, Ellie M. McCann, Sharon Powell, Sarah E. Cronin, and Veronica Deenanath. 2018. "Extension-Based Divorce Education: A Quasi-Experimental Design Study of the Parents Forever Program." *Journal of Divorce and Remarriage* 59(8):633–52. doi: 10.1080/10502556.2018.1466256.

Card, David, and Alan Krueger. 1994. "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania." *American Economic Review* 84(4):772–93.

Carnahan, Thomas, and Sam McFarland. 2007. "Revisiting the Stanford Prison Experiment: Could Participant Self-Selection Have Led to the Cruelty?" *Personality and Social Psychology Bulletin* 33(5):603–14. doi: 10.1177/0146167206292689.

Charlesworth, James E. G., Grace Petkovic, John M. Kelley, Monika Hunter, Igho Onakpoya, Nia Roberts, Franklin G. Miller, and Jeremy Howick. 2017. "Effects of Placebos without Deception Compared with No Treatment: A Systematic Review and Meta-Analysis." *Journal of Evidence-Based Medicine* 10(2):97–107. doi: 10.1111/jebm.12251.

Chetty, Raj, Nathaniel Hendren, and Lawrence F. Katz. 2016. "The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment." *American Economic Review* 106(4):855–902. doi: 10.1257/aer.20150572.

Correll, Shelley J., Stephen Benard, and In Paik. 2007. "Getting a Job: Is There a Motherhood Penalty?" *American Journal of Sociology* 112(5):1297–338. doi: 10.1086/511799.

Filmer, Deon, and Norbert Schady. 2009. "School Enrollment, Selection and Test Scores." Policy Research Working Paper No. 4998, World Bank, Washington, DC.

Fredrickson, Barbara L., Tomi-Ann Roberts, Stephanie M. Noll, Diane M. Quinn, and Jean M. Twenge. 1998. "That Swimsuit Becomes You: Sex Differences in Self-Objectification, Restrained Eating, and Math Performance." *Journal of Personality and Social Psychology* 75(1):269–84. doi: 10.1037/0022-3514.75.1.269.

Gaddis, S. Michael, and Nicholas DiRago. 2023. "Audit Studies of Housing Discrimination in the United States: Established, Emerging, and Future Research." SSRN. doi: 10.2139/ssrn.3796335.

Hagger, M. S., N. L. D. Chatzisarantis, H. Alberts, C. O. Anggono, C. Batailler, A. R. Birt, R. Brand, M. J. Brandt, G. Brewer, S. Bruyneel, D. P. Calvillo, W. K. Campbell, P. R. Cannon, M. Carlucci, N. P. Carruth, T. Cheung, A. Crowell, D. T. D. De Ridder, S. Dewitte, M. Elson, J. R. Evans, B. A. Fay, B. M. Fennis, A. Finley, Z. Francis, E. Heise, H. Hoemann, M. Inzlicht, S. L. Koole, L. Koppel, F. Kroese, F. Lange, K. Lau, B. P. Lynch, C. Martijn, H. Merckelbach, N. V. Mills, A. Michirev, A. Miyake, A. E. Mosser, M. Muise, D. Muller, M. Muzi, D. Nalis, R. Nurwanti, H. Otgaar, M. C. Philipp, P. Primoceri, K. Rentzsch, L. Ringos, C. Schlinkert, B. J. Schmeichel, S. F. Schoch, M. Schrama, A. Schütz, A. Stamos, G. Tinghög, J. Ullrich, M. vanDellen, S. Wimbarti, W. Wolff, C. Yusainy, O. Zerhouni, and M. Zwienenberg. 2016. "A Multilab Preregistered Replication of the Ego-Depletion Effect." *Perspectives on Psychological Science* 11(4):546–73. doi: 10.1177/1745691616652873.

Inzlicht, Michael. 2016. "Reckoning with the Past." *Michael Inzlicht* (blog). Retrieved June 9, 2023 (http://michaelinzlicht.com/getting-better/2016/2/29/reckoning-with-the-past).

Kirk, David S. 2009. “A Natural Experiment on Residential Change and Recidivism: Lessons from Hurricane Katrina.” *American Sociological Review* 74(3):484–505.

Kugelmass, Heather. 2016. “‘Sorry, I’m Not Accepting New Patients’: An Audit Study of Access to Mental Health Care.” *Journal of Health and Social Behavior* 57(2):1–16. doi: 10.1177/0022146516647098.

McCoy, Shannon K., and Brenda Major. 2003. “Group Identification Moderates Emotional Responses to Perceived Prejudice.” *Personality and Social Psychology Bulletin* 29(8):1005–17. doi: 10.1177/0146167203253466.

Meirick, Patrick C., and Stephanie Schartel Dunn. 2015. “Obama as Exemplar: Debate Exposure and Implicit and Explicit Racial Affect.” *Howard Journal of Communications* 26(1):57–73. doi: 10.1080/10646175.2014.986312.

Pager, Devah. 2003. “The Mark of a Criminal Record.” *American Journal of Sociology* 108(5):937–75. doi: 10.1086/374403.

Pager, Devah, Bart Bonikowski, and Bruce Western. 2009. “Discrimination in a Low-Wage Labor Market: A Field Experiment.” *American Sociological Review* 74(5):777–99. doi: 10.1177/000312240907400505.

Quillian, Lincoln, Devah Pager, Ole Hexel, and Arnfinn H. Midtbøen. 2017. “Meta-Analysis of Field Experiments Shows No Change in Racial Discrimination in Hiring over Time.” *Proceedings of the National Academy of Sciences* 114(41):10870–75. doi: 10.1073/pnas.1706255114.

Resnick, Brian. 2016. “What Psychology’s Crisis Means for the Future of Science.” *Vox*, March 14.

Resnick, Brian. 2018. “More Social Science Studies Just Failed to Replicate. Here’s Why This Is Good.” *Vox*, August 27.

Rodgers, Joseph Lee, Craig A. St. John, and Ronnie Coleman. 2005. “Did Fertility Go up after the Oklahoma City Bombing? An Analysis of Births in Metropolitan Counties in Oklahoma, 1990-1999.” *Demography* 42(4):675–92. doi: 10.1353/dem.2005.0034.

Rosenthal, Robert, and Lenore Jacobson. 1966. “Teachers’ Expectancies: Determinants of Pupils’ IQ Gains.” *Psychological Reports* 19(1):115–18. doi: 10.2466/pr0.1966.19.1.115.

Seelye, Katharine Q. 2018. “Devah Pager, Who Documented Race Bias in Job Market, Dies at 46.” *New York Times*, November 8.

Sherman, Lawrence W., and Richard A. Berk. 1984. “The Specific Deterrent Effects of Arrest for Domestic Assault.” *American Sociological Review* 49(2):261–72.

Stanford Center on Poverty and Inequality, dir. 2016. *Race, Employment and a Criminal Record: Devah Pager*.

Trochim, William M. K. 2020. “Knowledge Base.” Retrieved November 24, 2021 (https://conjointly.com/kb).

Vohs, Kathleen D., Brandon J. Schmeichel, Sophie Lohmann, Quentin F. Gronau, Anna J. Finley, Sarah E. Ainsworth, Jessica L. Alquist, Michael D. Baker, Ambra Brizi, Angelica Bunyi, Grant J. Butschek, Collier Campbell, Jonathan Capaldi, Chuting Cau, Heather Chambers, Nikos L. D. Chatzisarantis, Weston J. Christensen, Samuel L. Clay, Jessica Curtis, Valeria De Cristofaro, Kareena del Rosario, Katharina Diel, Yasemin Doğruol, Megan Doi, Tina L. Donaldson, Andreas B. Eder, Mia Ersoff, Julie R. Eyink, Angelica Falkenstein, Bob M. Fennis, Matthew B. Findley, Eli J. Finkel, Victoria Forgea, Malte Friese, Paul Fuglestad, Natasha E. Garcia-Willingham, Lea F. Geraedts, Will M. Gervais, Mauro Giacomantonio, Bryan Gibson, Karolin Gieseler, Justina Gineikiene, Elana M. Gloger, Carina M. Gobes, Maria Grande, Martin S. Hagger, Bethany Hartsell, Anthony D. Hermann, Jasper J. Hidding, Edward R. Hirt, Josh Hodge, Wilhelm Hofmann, Jennifer L. Howell, Robert D. Hutton, Michael Inzlicht, Lily James, Emily Johnson, Hannah L. Johnson, Sarah M. Joyce, Yannick Joye, Jan Helge Kaben, Lara K. Kammrath, Caitlin N. Kelly, Brian L. Kissell, Sander L. Koole, Anand Krishna, Christine Lam, Kelemen T. Lee, Nick Lee, Dana C. Leighton, David D. Loschelder, Heather M. Maranges, E. J. Masicampo, Kennedy Mazara,

Samantha McCarthy, Ian McGregor, Nicole L. Mead, Wendy B. Mendes, Carine Meslot, Nicholas M. Michalak, Marina Milyavskaya, Akira Miyake, Mehrad Moeini-Jazani, Mark Muraven, Erin Nakahara, Krishna Patel, John V. Petrocelli, Katja M. Pollak, Mindi M. Price, Haley J. Ramsey, Maximilian Rath, Jacob A. Robertson, Rachael Rockwell, Isabella F. Russ, Marco Salvati, Blair Saunders, Anne Scherer, Astrid Schütz, Kristin N. Schmitt, Suzanne C. Segerstrom, Benjamin Serenka, Konstantyn Sharpinskyi, Meaghan Shaw, Janelle Sherman, Yu Song, Nicholas Sosa, Kaitlyn Spillane, Julia Stapels, Alec J. Stinnett, Hannah R. Strawser, Kate Sweeny, Dominic Theodore, Karine Tonnu, Yasmijn van Oldenbeuving, Michelle R. vanDellen, Raiza C. Vergara, Jasmine S. Walker, Christian E. Waugh, Feline Weise, Kaitlyn M. Werner, Craig Wheeler, Rachel A. White, Aaron L. Wichman, Bradford J. Wiggins, Julian A. Wills, Janie H. Wilson, Eric-Jan Wagenmakers, and Dolores Albarracín. 2021. "A Multisite Preregistered Paradigmatic Test of the Ego-Depletion Effect." *Psychological Science* 32(10):1566–81. doi: 10.1177/0956797621989733.

Wolfers, Justin. 2015. "Why the New Research on Mobility Matters: An Economist's View." *New York Times*, May 4.

13. SURVEYS



In the 1970s, roughly one in five Americans reported on surveys that if their party nominated a qualified woman for president, they would not vote for her. A similar proportion felt the same way about voting for a qualified African American nominee. But attitudes changed over the decades, and in 2009, Barack Obama became the first African American U.S. president; in 2016, Hillary Clinton became the first woman presidential nominee of one of the two major U.S. political parties. By then, only a small fraction of Americans opposed voting for a woman or African American candidate. We know about this stark shift in voter opinion because of nationally representative surveys that have consistently posed this and other questions to U.S. adults over the past several decades. For example, the General Social Survey (GSS), a well-respected survey conducted every other year by the National Opinion Research Center (NORC), has included questions about bias in presidential voting since its first survey in 1972. (In 2010, the last time the GSS asked, only 3 percent and 4 percent of Americans said they would not vote for an African American or woman presidential nominee, respectively.)



Shirley Chisholm, the first African American woman elected to the U.S. Congress. announcing her historic candidacy for the Democratic presidential nomination in 1972. That year, 26 percent of Americans told the General Social Survey that they would not vote for a qualified woman candidate for president if their party nominated that person; 25 percent said the same thing about a qualified African American candidate. *Thomas J. O'Halloran, via U.S. Library of Congress*

Surveys like the GSS allow us to know with a considerable degree of precision how much the public's perspectives and preferences have evolved over the years. In *survey research*, a researcher poses a set of predetermined questions to a sample of individuals (**respondents**). Those questions (called **items**) are listed in a **questionnaire**—which a researcher can read off to a respondent during an in-person or remote interview, or which the respondent can answer on their own through a paper or online form. (Survey questionnaires are also called *survey instruments*, and all the defined procedures to implement a study, including the questionnaire, are collectively known as the *survey protocol*.) Typically, each item in a survey questionnaire offers two or more answers that the respondent can choose from (called **response categories** or *response options*), much like a multiple-choice exam.

Surveys are very much a part of our everyday lives. If you keep up with the news at all, you will inevitably hear the results of one kind of survey or another—from political polls that say who is the leading candidate in an election, to surveys of how much money people spend or what they think about particular issues. You've undoubtedly taken numerous surveys asking whether a customer service representative was polite and efficient in responding to your requests, or whether an instructor taught you what you wanted to know. Surveys are incredibly influential in virtually every political and economic context, providing key metrics for whether a policy gets enacted or whether a person gets promoted (or fired). Yet survey research is not always conducted in a rigorous fashion, which means that you should look at the findings from any surveys you encounter with a healthy degree of skepticism.

In this chapter, we'll talk about the strengths and weaknesses of survey research and the many types of surveys that sociologists conduct. We'll also discuss the considerations that go into devising effective survey questions and response options. As you'll learn, constructing a good survey is not an easy thing to do. It requires a great deal of thoughtful planning and often many rounds of revision.

Opening chapter image credit: *cottonbro studio, via Pexels. Adapted by Bizhan Khodabandeh.*

13.1. The Strengths and Weaknesses of Survey Research

Learning Objectives

1. Identify when it is appropriate to employ survey research for data collection.
2. Discuss the drawbacks of surveys as a research method.
3. Discuss the problems that might arise if your survey has a low response rate.

Although surveys are a quantitative research method, they have many of the same advantages of qualitative interviews. Like in-depth interviews, they are an excellent way to gather data on a phenomenon that is hard or impossible to directly observe. For instance, you could use surveys to understand people's preferences (such as their political orientation), their attitudes (such as their views toward immigrants), and their personality traits (such as their self-esteem). You could also get a sense of the level of understanding they have about a particular topic (such as what they know about a newly enacted law) and what behaviors or activities they engage in (such as whether they smoke and how much they drink).

The chief difference between surveys and in-depth interviews is that you sacrifice depth for breadth. The goal of surveys is not to gather rich detail about a few cases; it is to compare the responses of many, many individuals. In fact, of all the data-collection methods described in this book, survey research is probably the best method to use when one hopes to gain a *representative* picture of the attitudes and characteristics of a population that is too large to observe directly. Researchers can cover a large area—even an entire country—by using online, telephone, or mail-in surveys, and the results from their sample will be generalizable to the population of interest so long as they use rigorous probability sampling techniques (which we discussed in **Chapter 6: Sampling**). Surveys can employ probability sampling because they allow researchers to collect data from large samples for a relatively low cost. Indeed, if we consider the time, effort, and expense that goes into tracking down a single case for a given sample, surveys are more cost-effective than other research methods. For instance, a survey researcher could get a single respondent's overall take on an issue in mere minutes, whereas an in-depth interview would require much longer, given the open-ended questions that a qualitative interviewer would ask.

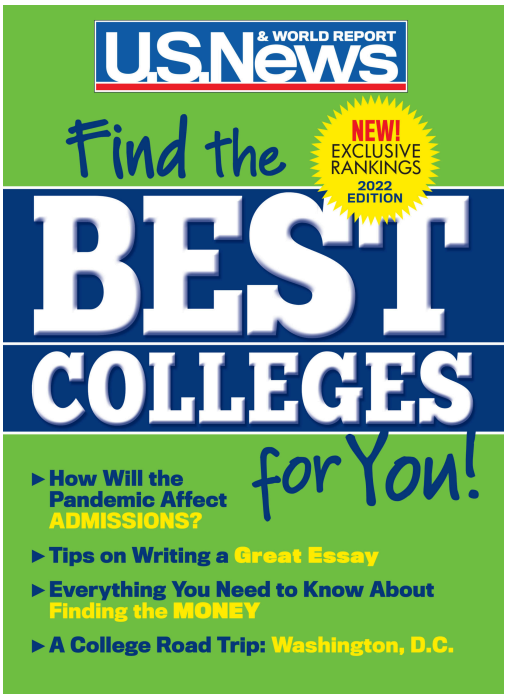
Survey research also tends to be a *reliable* method of inquiry (refer to **Chapter 7: Measuring the Social World**, for a discussion of reliability in research). This is because surveys are standardized: they pose the same questions, phrased in exactly the same way. This makes it possible for survey researchers to compare responses across the individuals in their sample in an apples-to-apples fashion. Other methods, including in-depth interviewing, do not achieve the same degree of consistency and comparability that a quantitative survey does. This is not to say that all surveys are always reliable. A poorly phrased question can cause respondents to interpret its meaning differently, leading to wildly inconsistent answers. If you construct your questionnaire carefully, however, it is likely to produce reliable results.

The *versatility* of survey research is yet another of its advantages. Knowing how to construct and administer surveys is a useful skill to have in all kinds of professions, and this knowledge is specialized enough that being proficient in survey design could land you a job all by itself. For instance, businesses use surveys to learn about their customers and ensure the quality of the services they provide. Governments use them to understand the opinions and needs of communities. Political campaigns and news outlets use them to capture public attitudes about issues and candidates. And all sorts

of other organizations—from social service agencies to churches and clubs to nonprofit and activist groups—regularly employ surveys to better understand their members or clients and evaluate the effectiveness of their efforts.

We should note, too, that surveys also do not have to use individuals as their **unit of analysis** (the type of cases that a researcher is studying). Most surveys are conducted at the individual level, but researchers also use this method to sample organizations, which may ultimately mean some representative of the organization fills out the form, but the details gathered relate to the group. One organizational survey you may have encountered is *U.S. News & World Report*'s "Best Colleges" rankings, which rely on surveys of college and university administrators who provide information about their own and other institutions.

Surveys can be used for exploratory, descriptive, or explanatory research. For instance, surveys can be used as a way to efficiently gather general details about your population of interest. Oftentimes, this description is all that nonacademic researchers really want to know: what the spending habits of their customers are, what their clients or voters believe is the most pressing issue, and so on. We discussed previously how in-depth interviews can be used to conduct **pilot interviews** in preparation for a larger study, but we should point out that a similar sort of exploratory research can be done with surveys. At the start of a research project, for instance, you might administer a quick survey that gives you a broad sense of the characteristics of your target population. This preliminary data collection could give you the necessary context to pursue more focused and time-intensive methods, perhaps helping you identify specific individuals or locations to examine with in-depth interviewing or ethnographic observation. (In fact, some surveys will ask respondents explicitly if they want to be contacted for follow-up interviews.)



While individuals must complete a survey, it does not necessarily have to collect data on an individual's characteristics. For example, to generate its influential annual rankings of colleges and universities, *U.S. News & World Report* relies in part on a survey of university administrators, who evaluate the academic quality of other institutions. Those ratings figure into the scores for each college and university. U.S. News & World Report, *via Wikipedia*

As for explanatory research, survey research is the method that many sociologists turn to when they want to engage in *causal inference*—that is, determining whether and to what extent a cause-effect relationship exists between two variables. As we discussed in **Chapter 4: Research Questions**, the explanatory research questions that survey research tends to tackle are focused on testing and measuring relationships between two variables, such as the following: *What effect does the independent variable have on the dependent variable*? Surveys are not as good as experiments are in ensuring **internal validity**—that a change in the independent variable truly does cause a change in the dependent variable—but they are often the only method available to sociologists when we want to study causal relationships in social life. With proper statistical controls (which we discuss later in the textbook), an analysis of survey data can do an adequate job of inferring causality. Meanwhile, the fact that most surveys collect data from a sample of the actual population of interest means that the findings from this research method tend to have a higher degree of **external validity**: we are more confident that any patterns and relationships we detect in our sample actually hold up in our larger population of interest (refer to **Chapter 12: Experiments** for a discussion of internal and external validity).

As with all methods of data collection, survey research also comes with drawbacks. First, surveys are *inflexible*—their questions cannot change after you begin your data collection (i.e., after you *field* your survey). In fact, altering a single word of a single item could potentially ruin the results you obtain from that question. Let's say you email a survey to

1,000 people. Within a few hours, the results start rolling in. Much to your horror, you realize that a number of your respondents are confused by the phrasing for a particular question. At this stage, however, it's too late for a quick fix. You can't just change the question for the respondents who haven't yet returned their surveys; if you did so, it would not be appropriate to compare their responses to those that the earlier group provided. The different wording means the two groups were essentially asked different questions, so you'd be comparing apples or oranges. You would need to resend the revised questionnaire to everyone who already filled out their forms to generate comparable data.

Consider how different this question-asking style is to the one adopted in in-depth interviews. Because comparisons between responses are not so crucial to the latter method, a qualitative researcher can tweak their questions—both between interviews and within a single interview. They can also pose entirely new questions that weren't on their interview guide, redirecting their line of inquiry as they learn more over the course of a conversation.

Validity can also be a problem with surveys. As we have noted, survey questions are standardized. They cannot change from respondent to respondent (with a few exceptions that we will get to later), and researchers cannot ask any follow-up questions that were not originally included as part of the questionnaire. As a result, it is difficult to ask anything other than general questions on a survey—questions that a broad range of people will understand. This means that the results you obtain from a survey may not be as valid as findings derived from methods like in-depth interviewing, which gives you the freedom to delve more deeply and comprehensively into whatever topic is being examined. Let's say that you want to learn something about voters' willingness to elect a nonwhite president, as in the chapter's opening example. GSS respondents were asked, "If your party nominated an African American for President, would you vote for him if he were qualified for the job?" Respondents could only respond "yes" or "no" to the question. But what if someone's opinion was more complex than a "yes" or "no" would indicate? What if, for example, a person was willing to vote for an African American woman but not an African American man? In that case, they might answer "yes" for the similar question on voting for a woman president, but none of the response options for the racial question would match their view. We are not suggesting that such a view makes any logical sense, but it is conceivable that an individual might hold it.

On surveys, a person with a complex (or convoluted) view might be forced to commit to simple yes/no answers (or another set of response options) that do not accurately reflect their actual point of view. As we will discuss later, survey questionnaires can address this problem by having catchall "Other" response options that a respondent can choose when their opinion doesn't align with the main response options. Nevertheless, the more fundamental problem of validity remains. The answers that respondents give on surveys must always be simplified for the survey to work its magic of being able to reduce a complex set of opinions into a broad pattern. Through this simplification, we can take Americans' complicated views on race and the presidency and boil them down to a single figure, such as 3 percent—the proportion of Americans who would not vote for a qualified African American presidential nominee in 2010, according to the GSS.

A final issue with surveys that we want to highlight is the potential for **nonresponse bias** (discussed in **Chapter 7: Measuring the Social World**), which can severely hamper this method's key strength of generalizability. Given how inundated people are nowadays with marketing requests of various kinds, researchers are finding it increasingly hard to get people to respond to their surveys. (See the sidebar **Response Rates and the Dirty Secret of Political Polling** to learn more about this worrisome trend.) As participation has dwindled, researchers have become increasingly concerned that the people who opt into their surveys look different from those who opt out in ways that might affect their findings. This danger—nonresponse bias—would mean that the samples they wind up with are not truly representative of the target population. People with particular characteristics have chosen not to join the sample, and if those characteristics have any bearing on the research question at hand—say, if Republicans happen to be less likely to answer our election poll (a real concern nowadays)—that will skew the results of the survey.

Response Rates and the Dirty Secret of Political Polling

A survey's **response rate** is the percentage of people who actually completed the survey out of everyone asked to participate. If you were handing out survey forms, the response rate would be the number of completed questionnaires you receive divided by the number of questionnaires you distributed. If you were calling people for a phone survey, it would be the number of calls that led to a person actually answering your survey questions divided by the number of all the phone numbers you dialed.

Let's say your sample included 100 people and you sent questionnaires to each of those people. It would be wonderful if all 100 returned completed questionnaires, but the chances of that happening—assuming you didn't put a gun to each potential respondent's head—are about zero. If you're lucky, perhaps 75 or so will return completed questionnaires. In this case, your response rate would be 75 percent (75 divided by 100). That's pretty darn good. Although response rates vary, and researchers don't always agree about what makes a good response rate, having three-quarters of your surveys returned would be considered good, even excellent, by most survey researchers.

Indeed, researchers who work on national opinion surveys and political polls often get much, much lower response rates than that. While not all pollsters publish their rates, those who do have produced alarming figures. For example, the Pew Research Center, a respected U.S. think tank that uses phone surveys, saw its response rate plummet from 36 percent in 1997 to just 6 percent in 2018 (Kennedy and Hartig 2019).[1] Pew researchers go to great lengths to eventually get the people they identify for their samples to agree to an interview, reaching out to each potential respondent *seven* times; other pollsters are far less rigorous.

Political polls are conducted regularly throughout the campaign season by a wide range of nonpartisan and partisan polling outfits—some more reputable than others. With the constant churn of new polls, researchers don't have the time to follow up, resulting in meager response rates. For instance, the *New York Times*/Siena College poll, a telephone survey, draws its phone numbers from the contact details that voters provide when registering to vote (notably, something that not all voters do). "In the poll we have in the field right now, only 0.4 percent of dials have yielded a completed interview," wrote *New York Times* chief political analyst Nate Cohn (2022) in October 2022, a month before the U.S. midterm election. Pollsters worry about low response rates because they fear it might be a sign of *nonresponse bias*, when respondents and nonrespondents differ in important ways. If only those who have strong opinions about your study topic return their questionnaires, for instance, the opinions in your sample would be more extreme than they would be if you had a truly representative sample of the target population.

1. For its phone surveys, Pew also provides information about its *contact rate* (households in which an adult was reached divided by all households called) and its *cooperation rate* (households that yielded an interview divided by all households in which an adult was reached). Those rates have also declined precipitously: between 1997 and 2018, Pew's contact rate fell from 90 percent to 62 percent, while its cooperation rate fell from 43 percent to 14 percent (Pew Research Center 2012).



Response rates for telephone surveys have fallen precipitously in recent decades, as people have become more reluctant to take calls from unknown numbers and—if they even pick up the phone—agree to participate. *Alex Green, via Pexels*

Some studies have concluded that rigorous efforts to boost response rates—namely, by following up with potential respondents and convincing them to participate—do not alter the findings of surveys in any substantial way (Curtin, Presser, and Singer 2000; Keeter et al. 2006). For their part, Pew researchers compared the characteristics of individuals who responded to one of their phone surveys to what they could learn through other means about those who didn't participate. (To learn about these nonresponders, they utilized two large national databases maintained by commercial vendors that have collected information on nearly every U.S. household.) Importantly, Pew found that there were no significant differences in voter registration or party affiliation between responders and nonresponders. However, responders were more likely to vote (54 percent versus 44 percent), more interested in community affairs and charities (43 percent versus 33 percent), and more interested in politics and current affairs (31 percent versus 25 percent). These results suggest that nonresponse bias will likely be a problem whenever we survey U.S. households about their civic and political engagement—given that people who are more engaged in these ways are also more likely to pick up the phone and answer survey questions—but it might not be a problem when we just want to know how the electorate will vote, which instead hinges on the mix of Democrats, Republicans, and third-party supporters in our sample.

Perhaps we don't have to worry about the abysmally low response rates of today's political polls, then. Although critics complain ("Polling seems to be irrevocably broken," a *Washington Post* columnist said after miscalls in

the 2020 U.S. presidential election), the industry continues to be as visible and influential as ever (Sullivan 2020). Even if they need to reach out to many more people nowadays to obtain adequate samples, pollsters continue to field surveys on a daily basis in the leadup to every major election.

Regardless of whether concerns about nonresponse bias in national surveys are overblown, we advise that you shoot for as high a response rate as possible for any surveys you conduct. Doing so will make your study's findings more credible. So, what is a good survey response rate? The truth is that there is no "magic number"—the response rate will vary depending on the method of data collection and on the target population. If you are targeting a known population, or if you are contacting a smaller number of people, your survey's response rate will likely be higher. Furthermore, some types of surveys are more effective than others at getting people to participate. One systematic review of the response rates in more than 2,000 healthcare surveys found that in-person surveys yielded the highest response rate (average of 76 percent), followed by mail surveys (65 percent), email surveys (51 percent), and web-based surveys (46 percent) (Meyer et al. 2022). (We discuss these different types of surveys in the next section.)

We suggest that you evaluate a survey's response rate in light of all these factors. Let's say you are sending mass emails to all of the students at a university. On the one hand, you might be able to get a list of all the students at that university (so in that sense, the population is well-defined—as opposed to, say, the U.S. population). On the other hand, the target population is quite large, and you are emailing your survey to students, so there's a good chance many people will simply ignore your requests. If you happen to receive a 50 percent response rate on this particular survey, that'd be a respectable number—around the average for an email survey. The same percentage would be low, however, for an in-person survey.

How do you avoid a low response rate? Here are some practical strategies to improve participation:

1. **Personalize your questionnaires:** If you are sending out your questionnaires by email or snail mail, address your recruitment message to specific potential respondents by name rather than to some generic recipient such as "Dear Resident" or "Dear Student."
2. **Highlight your credibility in describing your study:** In your recruitment pitch, provide ample details about the study and the credentials of any researchers involved, and include your full contact information so that potential participants can reach out with any questions. If at all possible, try to establish partnerships with respected institutions—such as university centers and departments, hospitals and clinics, and nonprofit organizations—so that respondents know you have institutional backing.
3. **Keep your questionnaire short and simple:** When potential respondents are deciding whether it's worth their time and effort to help you, the perceived length and complexity of your questionnaire will be a major factor. This is yet another reason to keep your survey instruments as short, simple, and focused as possible. Be sure to state upfront in your recruitment pitch that your survey only takes a few minutes to complete (if that is indeed the case).
4. **Be persistent in your recruitment efforts:** If you are asking a potential respondent over the phone or in person to participate in your study, be polite but persistent in your recruitment pitch. It's natural for people to hesitate about agreeing to a request from a stranger. One veteran political pollster told National Public Radio that the go-to line she uses in her recruitment pitch is, "How about we try some questions and see how it goes?" (Guo et al. 2022).
5. **Send out pre-questionnaire notices and post-questionnaire reminders.** A commonly used strategy to raise a survey's response rate is to send a pre-questionnaire notice. In this short message, you inform potential respondents that they will be asked to participate in a survey in the near future. This signals to the recipient that they should be on the lookout for the questionnaire and that the study being conducted is important (so important it needs to be announced beforehand!). After you send out the

actual questionnaire, you should send simple follow-up reminders to nonresponders after an appropriate interval of time (days or weeks, depending on the survey). If they still don't complete the questionnaire, don't be afraid to keep reminding them. This can only go so far, of course—not everyone will share your obsession with rigorous social science, and you don't want to infuriate or harass anyone—but you should make a good-faith effort to get people to respond. (Remember that the Pew Research Center contacts potential respondents *seven* times before giving up.) Note that the nudge you give can be very simple, too: a friendly reminder to those who have yet to complete the survey, paired with a thank you to those who have already returned it.

6. **Offer compensation or at least a token of appreciation:** To incentivize participation in surveys, researchers will often offer cash, gifts, or gift cards to those who complete their surveys. Even if you can't provide substantial compensation, a small gesture—such as including a $1 bill with every mailed questionnaire—can still increase responses. For an online survey, you can offer every respondent an opportunity to enter a raffle for one or more prizes (be sure to check applicable laws so that you do not run afoul of any regulations concerning raffles or rewards). Putting some money into purchasing, say, a tablet that you can give away just to one winner can boost your response rate substantially.

Key Takeaways

1. Survey research is often used by researchers who wish to describe or explain the characteristics of large groups. This is because it can employ probability sampling techniques that ensure a representative sample whose characteristics more or less match those of the target population.
2. In addition to its generalizability, the strengths of survey research include its cost-effectiveness, reliability, and versatility.
3. Although experiments are better at inferring causality, survey research can provide a good understanding of potential cause-effect relationships between two or more variables, allowing researchers to test whether a relationship exists and—if it does—measure the strength of that relationship.
4. Like qualitative interviewing, survey research can gather data about phenomena that the researcher cannot directly observe by allowing respondents to discuss things they have experienced or emotions or beliefs they currently have. Unlike questions in qualitative interviews, however, survey questions cannot be altered once data collection has begun, and this research method, by necessity, simplifies the opinions of respondents, which can pose problems of validity.
5. While survey researchers should always aim to obtain the highest response rate possible, some recent research argues that high return rates on surveys may be less important than we once thought.

Exercises

1. Recall some of the possible research questions you came up with while reading previous chapters of this textbook. How might you frame those questions so that they could be answered using survey research?
2. What are some ways that sociologists who use survey research might overcome the weaknesses of this method?
3. Find a journal article reporting results from survey research (refer to **Chapter 5: Research Design** for advice on searching for academic papers). How do the authors describe the strengths and weaknesses of their study? Are any of the strengths or weaknesses we've described here mentioned explicitly in that article?

13.2. Types of Surveys

Learning Objectives

1. Define cross-sectional surveys and outline some of the drawbacks of this approach.
2. Describe the various types of longitudinal surveys, and how retrospective surveys attempt to mimic longitudinal surveys.
3. Discuss the advantages and disadvantages of the various methods of administering survey questionnaires.

As a researcher, you can employ many different types of surveys, and they differ in two major ways: *how frequently* a survey is administered, and *how* it is administered to respondents. In this section, we'll go over these various approaches to conducting survey research.

Cross-Sectional versus Longitudinal Surveys

Cross-sectional surveys are surveys that are administered at just one point in time. You can think of cross-sectional surveys as a snapshot in time: they give us an idea about how things were for our respondents at the particular moment that they completed the survey.

One problem with cross-sectional surveys is that the opinions, behaviors, and other phenomena that such surveys are designed to assess don't generally remain the same over time. As a result, generalizing from a cross-sectional survey about the way things are can be tricky. Perhaps you can say something about the way things were in the moment that you administered your survey, but it is difficult to know whether things remained that way for very long afterward. Think about how Americans might have responded if they had been administered a survey asking for their opinions on terrorism on September 10, 2001. Now imagine how differently people would responded responses if they had been asked the same questions on September 12, 2001—the day after the attacks on the World Trade Center and Pentagon.

A more mundane example of this problem comes from preelection polling. (**Poll** is another word for "survey," though "poll**"** is more commonly used for nonacademic surveys that gauge voter preferences or public opinion.[1]) One of the goals of political polls is to give people a sense of who is "winning" or "losing" an election before the actual vote is completed and counted. Polls can be useful in tracking people's opinions over time in this way, but they are more

1. The word "poll" originally referred to the top of one's head, and its usage as a means of counting speaks to a time when political "polling" amounted to having people in a room move to different corners depending on which candidate they supported—with the number of "heads" in each corner tallied up for each candidate (Lepore 2015). A related term is straw poll, which refers to a nonscientific poll with no random sampling (which is pretty much useless, from a social scientific standpoint); according to the *New York Times* columnist William Safire, the term was meant to evoke the idea throwing hay into the air to see which way the wind was blowing (Brockell 2020). Refer to **Chapter 6: Sampling** for a discussion of how heavily political campaigns rely on polling to understand how potential voters think about candidates and what issues are important to them.

problematic when they are used as predictions about who will win an election. In addition to all the sampling and measurement problems that plague polling (some of which we discussed earlier), they are just a snapshot in time, and voter opinions—especially for that rare breed of undecided voters—can change rapidly between the moment that a poll is conducted and the moment that voters actually vote. According to some election analysts, this very kind of abrupt turnaround in voter sentiment happened right before the 2016 U.S. presidential election. Most mainstream political polls leading up to Election Day suggested that former secretary of state Hillary Clinton would win the presidency. However, some last-minute factors—such as damaging campaign revelations in the news—turned a crucial segment of voters against Clinton in the final hours of the campaign, which some pollsters (perhaps self-servingly) say is the main reason that she lost and their polls got it wrong.

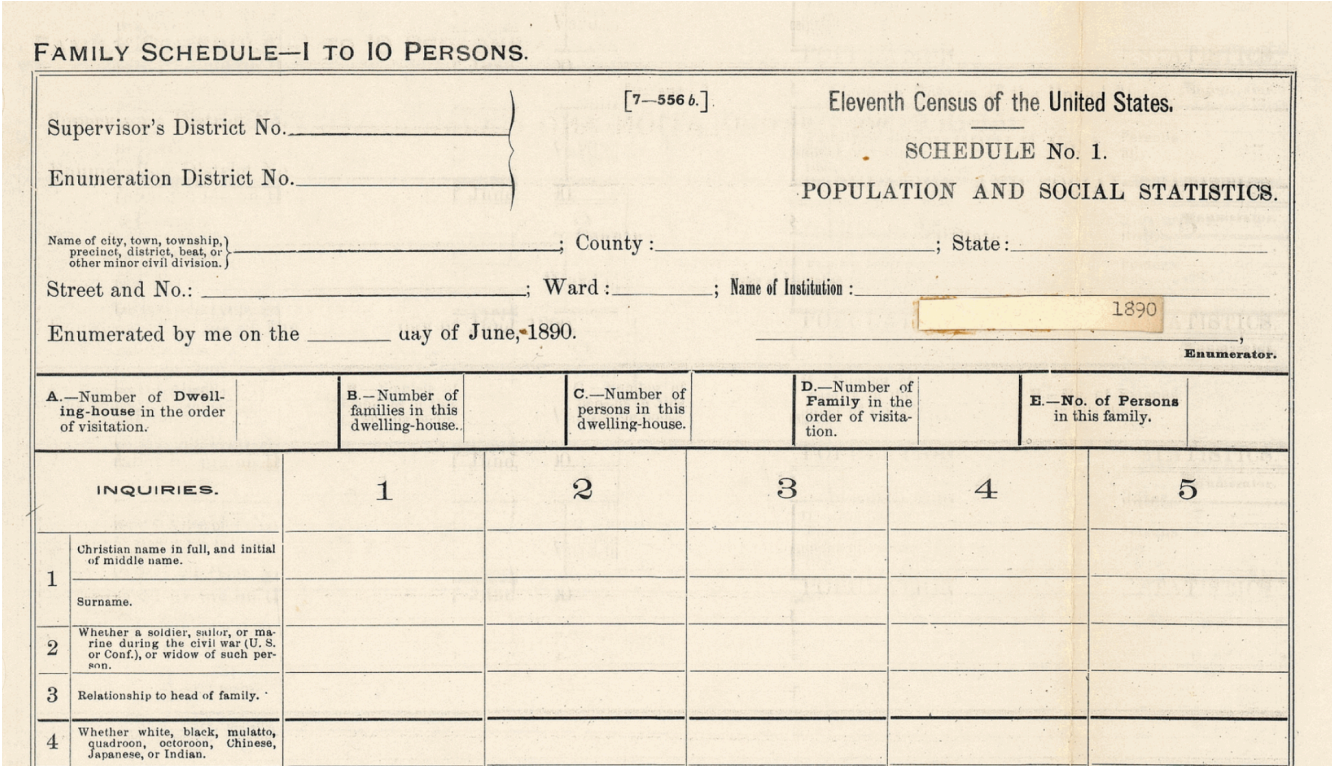
Longitudinal surveys make observations over some extended period of time. There are several types of longitudinal surveys, including *trend*, *panel*, and *cohort* surveys. Researchers conducting **trend surveys** are interested in how people's attitudes or behaviors change over time. A key characteristic of trend surveys is that they do not require interviewing the same people at each time point. Because the interest here is in trends, not specific people, as long as your sample is representative of whatever population you wish to describe over time, it isn't important that the same people participate each time.

The Gallup opinion polls are an excellent example of trend surveys (we discussed founder George Gallup's innovative approach to polling in **Chapter 6: Sampling**). To learn about how public opinion changes over time, Gallup administers the same questions to different groups of people at different points in time. For example, since 1936, Gallup has asked Americans, "Do you approve or disapprove of labor unions?" (Saad 2023). In 1936, 72 percent of Americans approved of unions. That figure peaked at 75 percent in 1953 and 1957 before declining in subsequent decades. Approval drifted below 60 percent in the 1980s, when U.S. president Ronald Reagan ushered in an era of anti-union politics and policies, and it reached its nadir in 2009—48 percent—as the belt-tightening of the Great Recession sharpened criticisms of relatively well-paid unionized workers in the public sector and elsewhere. In 2022 and 2023, however, approval of labor unions shot up to 71 and 67 percent, respectively, the highest levels seen since the 1960s. Perhaps not coincidentally, a wave of well-publicized strikes swept the country during these years.



In its well-known trend surveys, Gallup has asked Americans over the years about their approval of various institutions, including labor unions. Three-quarters of Americans approved of unions in the 1950s, but that support slackened over subsequent decades, reaching a low of 48 percent in 2009. In recent years, however, union approval has spiked, reaching 67 percent in 2023—the same year that high-profile strikes by Hollywood writers and actors and autoworkers dominated headlines. *Eden, Janine and Jim*, *via Flickr*

One problem with trend surveys is that the questions they use may have language that becomes outdated over time. As we have said, altering a question's wording in any way may alter how people respond to the question. As a result, it becomes impossible to know whether any dip or rise in the trend line following a wording change is due to the revised language or an actual movement in public opinion. Take the GSS question we started this chapter with: "If your party nominated an African American for President, would you vote for him if he were qualified for the job?" When the question was first asked in 1972, the word "Negro" was used; in later years, "Black" was used, and more recently, "African American." Clearly, the GSS needed to keep up with the times ("Negro" is so antiquated it would be jarring to hear in a survey interview), and you could argue that these minor alterations in wording did not lead people to answer the question in a substantially different way. The point is, though, we don't know for sure, and in some cases changing the wording in a questionnaire creates major problems for researchers. Take the questions about race and ethnicity that the U.S. Census Bureau has posed over the centuries. In the 1890 census, a question asked if the respondent was "white, black, mulatto, quadroon, octoroon, Chinese, Japanese, or Indian" (see **Figure 13.1**). In 1940, the questionnaire asked if the respondent was "White," "Negro," "Indian," "Chinese," "Japanese," "Filipino," "Hindu," or "Korean." Clearly, the categories for the 1890 census do not match those in 1940, much less those used today. These major changes in question wording over time make it hard to say with any precision what the actual numbers of people we would consider to be African American, Asian American, and so on today were in those earlier times when the census questions were very different. Sociologists Claude Fischer and Michael Hout tried to do exactly that in their book *Century of Difference* (2006), but this task required making many assumptions about the available data that social scientists generally prefer to avoid.



FAMILY SCHEDULE—1 TO 10 PERSONS.

[7—556 b.]

Supervisor's District No._____

Enumeration District No._____

Eleventh Census of the United States.

SCHEDULE No. 1.

POPULATION AND SOCIAL STATISTICS.

Name of city, town, township, precinct, district, beat, or other minor civil division. _____; County: _____; State: _____

Street and No.: _____; Ward: _____; Name of Institution: _____

Enumerated by me on the _____ day of June, 1890. 1890

_____, Enumerator.

A.—Number of Dwelling-house in the order of visitation.	B.—Number of families in this dwelling-house.	C.—Number of persons in this dwelling-house.	D.—Number of Family in the order of visitation.	E.—No. of Persons in this family.

	INQUIRIES.	1	2	3	4	5
1	Christian name in full, and initial of middle name.					
	Surname.					
2	Whether a soldier, sailor, or marine during the civil war (U. S. or Conf.), or widow of such person.					
3	Relationship to head of family.					
4	Whether white, black, mulatto, quadroon, octoroon, Chinese, Japanese, or Indian.					

Figure 13.1. 1890 Census Form. Question 4 in the 1890 U.S. census asked respondents about their household's racial and ethnic backgrounds. However, the response categories provided look nothing like the categories used in today's census, making it exceedingly difficult to compare the answers that year to those in subsequent decades, when the question wording and response options were different. *U.S. Census Bureau*

Even if the language a survey question uses is outdated, it sometimes makes sense not to change it so that responses in earlier years can be properly compared with those in recent years. For instance, since 1988, the GSS has occasionally asked respondents whether they agree or disagree with the statement "Being a housewife is just as fulfilling as working for pay." Nowadays, many people would use the term "stay-at-home parent" or something similar to convey the same idea, but the GSS has continued to ask this question in recent years without any edits to the wording. For trend surveys, seeing the ways that opinions evolve over time is the whole point of this longitudinal approach, and altering the question language could make the comparison more fraught in the ways we have described. Indeed, sometimes the relative changes over time in how people answer the same question are much more interesting than the breakdown of answers in a single year. The original question may not have been phrased in the best way, but whether its trend line goes up or down over time can still tell us a great deal.

In **panel surveys**, the same people participate in the survey each time it is administered. As you might imagine, panel studies are difficult and costly to conduct. Imagine trying to administer a survey to the same 100 people every year for, say, five years in a row. Keeping track of where people live, when they move, and when they die takes resources that researchers often don't have. Furthermore, panel studies present special problems of sampling bias because they almost always lose respondents over time: people die, or grow tired of the study, or drop out for other reasons. As we discussed in **Chapter 7: Measuring the Social World**, losing participants can generate a form of bias (**attrition bias**) because the people who remain may have certain characteristics—being healthier, more committed, more enamored with research, and so on—that can skew the study's findings in one or more directions.

When researchers successfully pull off a panel survey, however, the results can be quite powerful. Since 1988, researchers at the University of Minnesota have administered the Youth Development Study (YDS) every year to the same 1,000 people. Study participants were in ninth grade when the study began, and they are now middle-aged. Several hundred papers, articles, and books have been written using data from the YDS. One of the major lessons learned from this panel study is that doing paid work has a largely positive impact on young people (Mortimer 2005). Contrary to some popular beliefs, researchers found that work increases adolescents' confidence, boosts their academic performance, and prepares them for future careers. Without the unique data that this panel survey was able to obtain, researchers might not have been able to arrive at such a counterintuitive finding.

Another type of longitudinal survey is a **cohort survey**. In a cohort survey, a researcher identifies some category of people who are of interest and then regularly surveys people who fall into that category. The same people don't necessarily participate from year to year, but all participants must meet whatever **inclusion criteria** the researcher has set to be recruited for the sample. (Note that it is possible to have *panel cohort* surveys, and the YDS is an example of one: the respondents were all ninth-graders when the study began.)

The sorts of cohorts that researchers frequently study include people born around the same time period (i.e., generations), graduating classes, people who began work in a given industry at the same time, and people who have some specific life experience in common. A good example of an analysis of generational cohorts is a study by Christine Percheski (2008) of women's employment. Percheski compared employment rates for women born across seven different generations: from Progressives born between 1906 and 1915, to Generation Xers born between 1966 and 1975. One of her findings was that the labor force participation rate—the percentage of the population either employed or actively seeking jobs—increased consistently across these cohorts of women. Percheski also found that Gen X professional women with young children had higher labor force participation than similar women from previous generations. Her analysis contradicted the speculation of some journalists that Gen X mothers were opting out of the workforce en masse (Belkin 2003).

One substantial benefit of conducting panel studies is that they make it easier to infer causality—that is, evaluate whether a cause-effect relationship exists between two variables. This is a research goal that is near and dear to the hearts of many sociologists, who don't want to just describe reality, but explain it. As we discussed in **Chapter 12: Experiments**, experiments excel at this task because they randomly assign subjects into control and experimental groups and then introduce a treatment or stimulus. This research design makes it easy to infer that a change in the independent variable truly caused any outcomes that occurred in the experimental group. Among other things, the fact that the subjects were randomly assigned to the two groups rules out the possibility that preexisting differences between those groups explain those observed outcomes.

But with a cross-sectional survey, you have only observational data, not experimental data. As we'll discuss in **Chapter 14: Quantitative Data Analysis**, you need to apply statistical procedures in a cross-sectional analysis to ensure that any outcomes you observe are the result of changes in the independent variable rather than preexisting differences. But in a panel study, you essentially have built-in control and experimental groups: the research participant earlier in time is the control group, and the participant later in time (after the change in the independent variable) is the experimental group. This allows you to say with greater certainty that any changes in outcomes that a participant experienced were driven by changes in the independent variable of interest.

Let's say you are conducting a study of the impact that having a criminal record has on a person's employment prospects. If you had access to cross-sectional survey data, you would likely see that individuals with criminal records are much less likely to be employed. You wouldn't know, however, if this association was due to the criminal record, or some confounding variable—for instance, the fact that individuals with criminal records also tend to have personality traits that make them less employable. With panel data, however, you might be able to follow a particular individual before and after they were arrested. If you saw that the individual's ability to obtain and maintain employment worsened after their entanglement with the legal system, you would have stronger evidence that a causal relationship truly exists between

criminal records and poor employment outcomes, based on an assumption that factors like personality traits would have stayed more or less constant across the pre-arrest and post-arrest scenarios.

As we noted in **Chapter 12: Experiments**, you would have even stronger causal evidence if you conducted an experiment—as Devah Pager (2003) did with her groundbreaking audit study to measure the "mark" of a criminal record—but field experiments of this kind are also exceedingly hard to implement. A well-done panel study can tell you a great deal about causal relationships, and they have the added benefit of greater external validity—giving you a richer sense of how a phenomenon unfolds in the real world.

Consider the panel study that David Harding, Jeffrey Morenoff, and Jessica Wyse conducted for their book *On the Outside* (2019), which tracked Michigan state prisoners released on parole. Their study included a quantitative analysis of **administrative data**—data from an organization's own records (in this case, the Michigan state government). Rather than needing to administer their own surveys, the researchers had access to Michigan's data on every prisoner released in 2003—11,000 individuals, or the *entire* target population—whom they were able to track in state records from the date of their parole until seven years later. To flesh out the overall trends in their quantitative data, the researchers also conducted in-depth interviews and observations of 22 of those parolees, the study's *focal participants*, over that time period. In all these ways, the three sociologists took advantage of their study's robust longitudinal design to examine the factors that make it difficult for parolees to reintegrate into society after their release from prison.

Like cross-sectional surveys, **retrospective surveys** are administered only once, but they are still able to capture changes over time by asking respondents to report past behaviors, beliefs, or experiences. By relying upon respondents' recall, retrospective surveys gather longitudinal-like data without actually incurring the time or expense of a real longitudinal survey. Of course, this benefit must be weighed against the very real possibility that people's recollections are faulty. Imagine, for example, that you're asked in a survey to respond to questions about where, how, and with whom you spent last Valentine's Day. Given that the last Valentine's Day can't have been more than 12 months ago, chances are good that you can respond accurately to any survey questions about it. But now let's say the researchers want to know how the last Valentine's Day compares with previous Valentine's Days, and so they ask you to report on the preceding six Valentine's Days as well. How likely is it that you will remember details from that far back? Clearly, your retrospective responses will not be as accurate as they might have been had you been asked that same question each year over the past six years—that is, as part of a panel survey.

We've discussed cross-sectional and longitudinal research designs in relation to survey research because these distinctions are made most commonly between different types of surveys. However, we should note that surveys are not the only method of data collection that can be cross-sectional or longitudinal. For instance, you could conduct multiple ethnographic observations of the same place over a certain span of time, or you could conduct in-depth interviews with a similar cohort of respondents every few years. You could even use multiple methods in a panel study: consider the panel study of parolees we just mentioned that relied on both qualitative and quantitative data. In fact, if the idea of doing a panel study has piqued your interest, you might want to check out the *Up* series of documentary films. While not academic research, the films offer an artistic approximation of a panel study. Filmmakers began filming the lives of 14 British children in 1964, when the children were seven years old. They have since caught up with the children every seven years. The ninth and most recent installment of the documentary, *63 Up*, premiered in 2019 (see **Video 13.1**).



One or more interactive elements has been excluded from this version of the text. You can view them online here: https://viva.pressbooks.pub/sociology-research-methods/?p=447#oembed-1

Video 13.1. The *Up* Documentary Series. Although the documentaries in Michael Apted's *Up* series are not research studies, they do illustrate the general approach of panel surveys, which follow the same people over a particular time period—often several years, and sometimes for decades. In *7 Up* (1968), Apted profiled 10 boys and 4 girls from varied class backgrounds when they were seven years old. Every seven years, he released a new documentary chronicling their lives in the years since, with the most recent installment, *63 Up* (2019), featuring 11 of the 14 participants (one died and two did not participate).

Self-Administered versus Interview-Administered Questionnaires

Surveys can be grouped into two categories based on how their questions are asked and answered. The first is **self-administered questionnaires**, which present respondents with a set of questions that they answer on their own. The second is **interviewer-administered questionnaires**, which entail having a researcher pose each question aloud and then record the respondent's answer. On the one hand, interviewer-administered questionnaires involve much more effort on the researcher's part than self-administered questionnaires, and they present risks of certain biases because the respondent is answering questions in the presence of another person. On the other hand, having a researcher oversee the completion of the questionnaire helps ensure that respondents take the survey seriously and answer questions carefully and with greater clarity about what is being asked. We will discuss each approach in turn.

There are three ways of delivering self-administered questionnaires: in person, via mail, or online. In the first approach, researchers hand out self-administered questionnaires to the study's participants in person. Traditionally, the in-person approach has involved handing out printed-out copies of the survey questionnaire, but increasingly, researchers working on site will let respondents borrow tablets or laptops or use their own phones to access an online version of the questionnaire (thereby automating the data-entry process, as we will discuss).

Frequently, an in-person survey will be given to an entire group at one time. These *group-administered surveys* are very common on college campuses (researchers love to administer surveys in large lecture classes, where they have a more or less captive audience), but they can be employed in virtually any setting. In this approach, a sample of respondents is brought together at a common place and time, and each respondent is asked to complete the survey questionnaire while in that room. Respondents enter their responses independently without interacting with each other.

Group-administered surveys have a number of advantages. They ensure a high response rate because many of those asked to fill out the survey in such a setting will do so. If respondents do not understand any specific question, they can ask for clarification. And in many organizations, it is relatively easy to assemble a group of people in a conference room or cafeteria, especially if the survey is approved by higher-ups.

If you are not able to visit each member of your sample personally to deliver a survey, you might consider sending your survey through the mail. In this case, the same questionnaire is sent out to a large number of people. Respondents can complete the survey at their convenience and return it in the (hopefully postage-prepaid) envelopes you provide. The key advantages of mail surveys are that they are unobtrusive and inexpensive to administer (bulk postage is cheap in most countries). Furthermore, you can draw upon a sampling frame of all the postal addresses in a given geographical area, whereas other means of administering surveys do not always have an obvious sampling frame. (For instance, there is no list of every American's email address—at least not any a sociologist is privy to.) However, you will need to mail a lot of surveys just to get a few responses because most people tend to ignore survey requests sent to them in the mail—or anything sent in the mail these days, to be fair. Even if they intend to fill out the survey, many respondents may wind up

losing it. There may also be long delays of several months between when the surveys are sent out and when researchers get them back, completed, in the mail.

Because of these drawbacks, many researchers use mail surveys in conjunction with other methods. For instance, they may mail a survey that also contains a link to the online version of it, so that the respondent can fill out the questionnaire in whatever way they prefer. Oftentimes survey researchers will deliver the survey or an announcement about it (**Figure 13.2**) through snail mail, and then follow up with a phone call or email, thereby providing multiple reminders about the survey (and possibly multiple ways to complete it).

2023 Survey of Doctorate Recipients

September 2023

Dr. Victor Chen
Vcu
827 W Franklin St
Richmond, VA 23284-9003

Dear Dr. Chen,

You are invited to complete the **Survey of Doctorate Recipients (SDR)** sponsored by the National Science Foundation and the National Institutes of Health. Thank you for your past contribution to this unique study of doctorate holders. Since 1973, the SDR has studied the experiences of doctorate degree holders and is the only source of data on the careers of science, engineering, and health doctorate holders from U.S. academic institutions.

To complete the SDR online, open the survey link on a computer or any mobile device and enter your unique personal identification number (PIN):

SDR SURVEY LINK:	**https://www.NSFsdr.org**
PIN:	[redacted]

Your continued participation in this research will help ensure the validity and accuracy of the survey data. SDR results provide crucial information about the educational and career outcomes of the most highly educated segment of our population.

The information you provide will be collected by NORC at the University of Chicago, the survey contractor conducting the SDR on our behalf. Your responses are confidential and will only be used for statistical purposes. If you have any questions regarding the survey, please review the information on the back of this letter or contact NORC by emailing SDR@norc.org, calling 1-800-685-1663 (toll-free in the U.S. and Canada) or 312-871-4272, or visiting https://www.NSFsdr.org. NORC staff are available from 9 a.m. to 9 p.m. (Central Time) to assist you.

Thank you for your continued contribution to the SDR.

Sincerely,

Emilda B. Rivers, Director
National Center for Science and Engineering Statistics
National Science Foundation

P.S. To ensure this message reaches you, we are also sending this survey invitation to the email address NORC has on record for you.

INVLTR_2_M_01_11103450

For assistance completing the survey:
Call 1-800-685-1663, email SDR@norc.org, or visit https://www.NSFsdr.org

For more information about the survey:
Visit https://ncses.nsf.gov/surveys/doctorate-recipients or write NSF at 2415 Eisenhower Avenue, Alexandria, VA 22314

0000042220

Figure 13.2. Letter with a Survey Link. This letter invites participation in a survey of U.S. doctorate recipients sponsored by the National Science Foundation and National Institutes of Health. While the survey questionnaire is typically completed online, the government agencies overseeing the survey pursue a range of strategies to ensure a high response rate: emailed invitations to complete the survey, snail-mailed letters (like this example) with a link to the online survey, multiple follow-up phone calls and texts, and even gift cards mailed to entice particularly recalcitrant respondents.

Creating online questionnaires has become easy to do through platforms like SurveyMonkey, Google Forms, and RedCap, which administer surveys over the internet using interactive forms. This approach is relatively cheap (many platforms are free, at least for limited usage), and online forms can gather and process your data much more quickly than other methods of administering surveys. Furthermore, any changes you make to the survey occur instantaneously, without the need to print out another batch of forms—but remember that you should not change the questions once a survey has been fielded.

To disseminate their online surveys, researchers can post a link to the survey on email listservs, social media platforms, discussion boards, or listings sites. They can target specific individuals with emails containing the survey link. They can even include a survey link on any physical flyers they distribute—perhaps using QR codes to encode the link for easy retrieval by potential respondents. Note that the rising popularity of online surveys has also created a cottage industry of for-hire survey respondents on platforms like Amazon's Mechanical Turk or Qualtrics; some researchers who have found it hard to recruit respondents for their studies have posted surveys there, which freelancers fill out in exchange for compensation. (Buyer beware: in addition to the difficulties that online surveys inevitably face in gathering a representative sample, there are particular concerns about the quality of the responses obtained through these freelancer sites.)

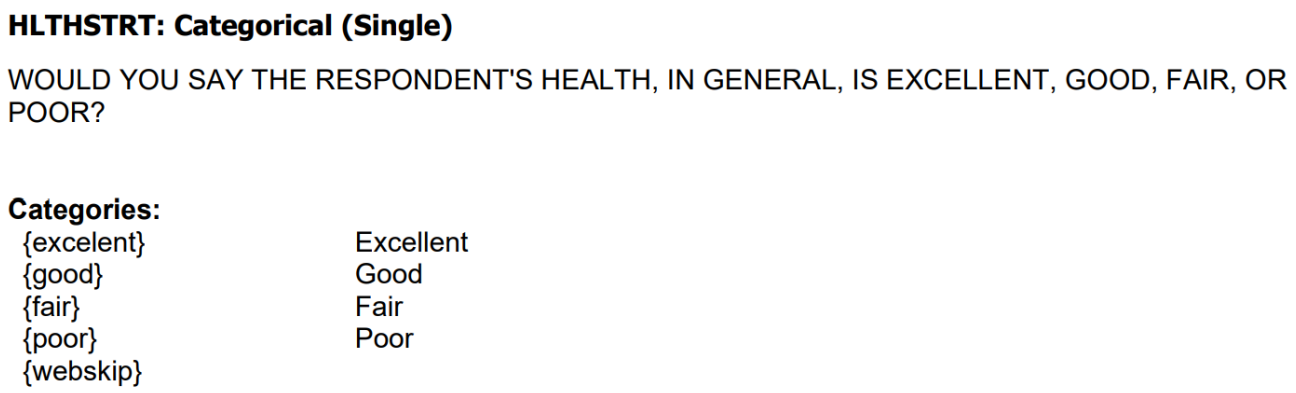
One major advantage of using online survey platforms like SurveyMonkey and Google Forms is that they make the job of data entry much easier. Rather than having to take all the completed hard-copy questionnaires and tally up answers on every survey item, you can rely on the platform to do that work for you. Furthermore, some platforms allow you to download your survey results into file formats that are readable by data analysis programs such as SPSS and Microsoft Excel (see **Chapter 14: Quantitative Data Analysis** for a discussion of these programs). That means you are not restricted to doing your data analysis using the platform's built-in tools, which can be quite rudimentary.

A key drawback of online surveys is security. With paper forms, a promise of anonymity may be easier to keep: once the data from the completed questionnaire has been harvested, the form can be destroyed, making it harder to link the answers back to the respondent. With all the identifying details that our online actions generate, however, you will need to be extra careful about the data you collect in this way; at the very least, you need to store your data someplace where it is password-protected and otherwise secure. Online surveys also face unique challenges, such as the need to prevent multiple submissions from the same user or to keep bots or trolls from sabotaging a publicly available survey. Computing the response rate of these surveys may also be difficult to do if the survey link is posted on email listservs or online platforms instead of being emailed directly to targeted respondents. And while every kind of survey must deal with problems of sampling bias, online surveys are particularly vulnerable to it. As noted earlier, they do not lend themselves to an obvious sampling frame, which means that researchers who use online surveys often have to use quota sampling and weight their sample to match what they know about the target population (see our related discussion in **Chapter 6: Sampling**). Another issue that online surveys deal with is the so-called digital divide: they cannot reach people who do not have internet access or who are not sufficiently tech-savvy to complete an online survey. Even in rich countries like the United States, many rural areas and other marginalized communities lack decent and affordable internet access. Without proper adjustments, online surveys are likely to obtain samples that are skewed toward a younger, more urban, and more affluent demographic.

The alternative to self-administered questionnaires is having a researcher walk the respondent through each of the survey's questions. When an *interviewer-administered questionnaire* is being fielded, the researcher conducts an

interview with each respondent, posing questions directly to them rather than having them read the questions on their own. Unlike an in-depth interview, however, the conversation is very much scripted: the interviewer reads each question in the questionnaire word for word, without any changes or improvised follow-up questions.

Interviewer-administered surveys are a more personalized form of data collection method than self-administered surveys. Importantly, the interviewer has the opportunity to clarify any issues raised by the respondent as the two of them go through the survey questions, hopefully addressing any confusion or concerns that may prevent the respondent from answering to the best of their ability. While the interviewers are required to follow a standardized set of questions, the interview script may contain special instructions for the interviewer that are not seen by respondents, which can affect the ordering of questions and other aspects of the interview. The interviewer may also be tasked with recording their personal observations and reflections during the course of the interview, and as part of the survey protocol, they may even be presented with specific questions to answer about the respondent, such as generally assessing that person's health—perhaps in addition to asking the respondent to assess their own health (see **Figure 13.3** for an example).



HLTHSTRT: Categorical (Single)

WOULD YOU SAY THE RESPONDENT'S HEALTH, IN GENERAL, IS EXCELLENT, GOOD, FAIR, OR POOR?

Categories:

{excelent}	Excellent
{good}	Good
{fair}	Fair
{poor}	Poor
{webskip}	

Figure 13.3. Questions for Interviewers. A question from the General Social Survey's 2022 questionnaire that asks the interviewer to rate the respondent's health. *NORC*

In addition to demanding substantially more time and effort from researchers than self-administered questionnaires do, survey interviews also require trained interviewers. The researcher working in the field must be personable and enthusiastic enough to convince the respondent to sit through all the questions on the questionnaire, and they must be disciplined enough not to deviate from the survey protocol in any substantial way—that is, in any way that would make it harder to compare one respondent's answers with another's. The interviewer is also considered to be part of the measurement instrument, in the sense that their presence can alter the behavior of the respondent (what sociologists call **interviewer effects**). Survey researchers must consider how the identity of the interviewer (e.g., their age, gender, race) might influence how the respondent answers certain questions, and whether they need to vary those characteristics across their interviewers to diminish any bias created by these interviewer effects. They might also need to train the interviewer to avoid acting in ways that might bias the respondent's answers—for instance, conveying emotions of disgust or excitement through their facial expressions in response to certain questions or response options. In general, **social desirability bias** is a greater problem with survey interviews, given that respondents can answer self-administered questionnaires without having to divulge any controversial opinions or sensitive details directly to another human being.

Two common types of survey interviews are *face-to-face interviews* and *telephone interviews*. In a face-to-face interview (also known as a *personal interview*), the interviewer meets the respondent in person. This is an approach taken by

some major surveys (including the U.S. census), which may start with a sampling frame of physical addresses and send out interviewers to a sample of those addresses. As we noted earlier, face-to-face interviews tend to have the highest response rates of all survey methods: people find it harder to tell a person to their face that they do not want to participate or they want to drop out. (That said, some respondents will be more uncomfortable being interviewed in person, depending on their personality and the nature of the questions.) Particularly skilled interviewers can persuade even the most reluctant individuals to cooperate, dramatically improving response rates.

In telephone surveys, interviewers contact potential respondents over the phone, typically using some sort of directory or a random range of phone numbers as a sampling frame. The most technologically sophisticated approach to this type of survey research is **computer-assisted telephone interviewing** (CATI), which is widely used by academic, government, and commercial survey researchers. In this setup, the interviewer is guided through the interview process by a computer program that automatically pulls up instructions and questions to be asked on the interviewer's screen. The system also selects respondents randomly using a random digit-dialing technique, and it records their responses using voice-capture technology.

Like with online surveys, telephone interviewing is not ideal for rural areas, where phone density is low. It also cannot be used to communicate non-audio information, such as images or video that a survey researcher wants respondents to react to. And in part because of the widespread use of caller ID, response rates for phone surveys have plummeted: many people no longer bother to answer calls from unknown phone numbers. If researchers can eventually get a respondent on the phone (through repeated calls and voicemails), they have a decent chance of getting them to agree to be interviewed. Remember our earlier advice: be polite but persistent, and ease the respondent into an interview with a line like, "How about we try some questions and see how it goes?"

Key Takeaways

1. Cross-sectional surveys are administered at one time, providing a single snapshot of social reality. By contrast, longitudinal surveys are administered over time, allowing researchers to track how attitudes evolve and observe cause-effect relationships more directly.
2. Retrospective surveys offer some of the benefits of longitudinal research but also come with a serious drawback—that is, the lack of reliability in a respondent's ability to recall the past.
3. Self-administered questionnaires can be delivered in a hard-copy form to participants in person or through snail mail, but they are increasingly administered through online forms.
4. Interviewer-administered questionnaires allow a researcher to clarify questions, make observational notes, and ensure that the respondent is taking the survey seriously, but they also raise concerns that the interviewer's identity or actions might influence how the respondent responds.

13.3. Designing Effective Questions and Questionnaires

Learning Objectives

1. Describe some of the ways that survey questions might confuse respondents and how to address these potential problems.
2. Identify some of the guidelines you should follow when devising response options.
3. Discuss why pretesting is important.

In this section, we'll provide some tips on posing clear survey questions that yield useful data. We'll also give advice on how you might present those questions in your questionnaire. As you'll see, constructing a survey questionnaire is an art. You must make numerous decisions—not just about the content of your questions, but also their wording, format, and sequencing. All of these factors can have important implications for how people respond to your survey (see **Video 13.2**).

Asking Effective Questions

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://viva.pressbooks.pub/sociology-research-methods/?p=455#oembed-1

Video 13.2. Writing Survey Questions. This video from the Pew Research Center goes over some pointers for posing good questions.

Much of the advice that we gave about devising questions in **Chapter 10: In-Depth Interviewing** applies to survey questions, too. Avoid biased questions. Steer clear of jargon and be very clear about what you are asking. The key difference, of course, is that the questions for a typical survey are entirely or almost entirely closed-ended. Furthermore, the inability to ask follow-up questions means that your questionnaire must be comprehensive, leaving nothing important out. And because you are posing the exact same questions word for word to every respondent, your standards for the clarity and precision of your language need to be even greater. You'll definitely want to pretest your questions before fielding the questionnaire, as we will discuss later. But let's start by reviewing some fundamentals, which will necessarily repeat some of our advice in earlier chapters.

The first thing you need to do to write effective survey questions is identify what exactly it is that you wish to know. Let's say you want to understand how students at your school made the transition from high school to college. Perhaps you wish to identify which students were more or less successful in this transition and which factors contributed to

their success or lack thereof. To get at these research questions, your questionnaire will need to find some justifiable way of operationalizing "success" (see our discussion of operationalization in **Chapter 7: Measuring the Social World**). Specifically, you should decide which questions you should ask to measure the "success" of someone's high school-to-college transition. Your questionnaire will also need to ask about all the factors that might have influenced each student's likelihood of success in that transition. Consulting the literature on the topic will certainly help, but you should also take the time to do some brainstorming on your own and to talk with others (perhaps even conducting exploratory in-depth interviews) about what they think is important in the transition to college.

Once you've identified all the relevant topics you want to cover in your questionnaire, you'll need to write specific questions on those topics. Make those questions as clear and to the point as possible. This is not the time to show off your poetic writing style; a survey is a technical instrument and should be written in a way that is as direct and succinct as possible.

Although we have stressed the importance of being comprehensive, at the same time, you don't want to take an everything-but-the-kitchen-sink approach by uncritically including every possible question that occurs to you. Doing so puts an unnecessary burden on your survey respondents. To help you trim a long questionnaire, rank your questions and include only those that you view as most important. Otherwise, your respondents might give up and neglect to fill out the entire form—or, even worse, they might start giving random answers just to end their suffering, leaving you with garbage data. (Partly to address these problems, larger surveys like the GSS use a **split-ballot design** whereby the sample is split into groups, and each group receives a different set of supplementary questions in addition to a universal set of questions that everyone answers; this strategy of having multiple questionnaires, or *ballots*, increases the total number of questions covered while not further lengthening the survey interview.)

Make sure that every question you pose will be relevant to every person you ask to answer it. This means two things: (1) your respondents should have adequate knowledge about whatever topic you are asking them about; and (2) they should have experience with whatever events, behaviors, or feelings you are asking them to report. For example, you probably wouldn't want to ask a sample of 18-year-old respondents how they would have advised U.S. president Ronald Reagan when news of his administration's sale of weapons to Iran broke in the mid-1980s. For one thing, few 18-year-olds are likely to have any clue about how to advise a president. Furthermore, the 18-year-olds of today were not even alive during Reagan's presidency, so they have had no experience with the event you would be asking them about.

Let's apply this criterion of relevance to our example study examining students' transitions from high school to college. In this case, our respondents should understand what exactly we mean by "transition to college" if we explicitly use that phrase in our questionnaire; otherwise, we need to break down that jargon into simpler and more concrete terms. Furthermore, we probably want to find respondents who have actually experienced the transition to college if we are going to ask them about that topic (unless understanding younger people's expectations about that transition—which may or may not turn out to be realistic—is our goal, of course).

Even more so than in-in-depth interviewing, the responses of survey respondents are very sensitive to the types of questions asked. You cannot follow up to tease out what people actually think, so poorly framed or ambiguous questions in a survey will likely result in meaningless responses with little value. Survey researchers should avoid questions that suffer from the following problems, all of which are likely to confuse at least some of your respondents:

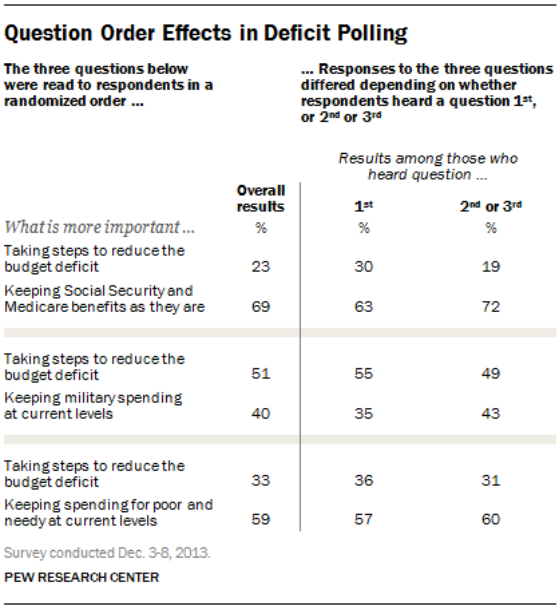
1. **Jargon, culturally specific terms, and other potentially confusing phrasings:** Survey questions should be stated using simple language, preferably in active voice, and without complicated words or jargon that may not be understood by a typical respondent. All questions in the questionnaire should be worded in a similar manner to make it easy for respondents to read and understand them. You should also avoid using terms or phrases that may be regionally or culturally specific. The exception, of course, is if all your respondents come from the region or culture whose terms you are using. Likewise, if your survey is targeted at a specialized group of respondents, such as doctors, lawyers, and researchers, using jargon may make sense. These respondents would be using such terms

in their everyday environment, and you don't want to bog down your questionnaire with unnecessary explanations of terms.

2. **Double negatives:** Any time respondents are forced to decipher questions that utilize two forms of negation, confusion is bound to ensue. Consider this question: "Did you not drink during your first semester of college?" A response of "no" would mean that the respondent *did* actually drink—they did not *not* drink. Generally speaking, avoiding negative terms in the wording of your questions will make their meaning clearer.[1] An infamous example of the bewilderment created by negative terms is a 1992 Roper poll that asked, "Does it seem possible or does it seem impossible to you that the Nazi extermination of the Jews never happened?" A shocking 22 percent of respondents said it was "possible" the Holocaust never happened, which greatly alarmed the group that had commissioned the poll, the American Jewish Committee. However, when Roper fielded another poll two years later, it reworded the question to make its meaning clearer: "Does it seem possible to you that the Nazi extermination of the Jews never happened, or do you feel certain that it happened?" This time, only 1 percent of respondents said it was possible that the Holocaust had never happened—still a worrisome figure, but one that didn't portend an extreme and sudden spike in Holocaust denialism (Kagay 1994).
3. **Double-barreled questions:** When we ask **double-barreled questions**, we ask multiple questions as though they are a single question. This is a problem because survey respondents may not know how to respond. Let's say we asked the following question in our survey about the transition to college: "Did you find the classes you took during your first semester of college to be more demanding and interesting than your high school classes?" How would someone respond if they felt their college classes were more demanding but also more boring than their high school classes? Or less demanding but more interesting? Because the question combines "demanding" and "interesting," there is no way to respond "yes" to one criterion but "no" to the other.
4. **Leading or otherwise biased questions:** As we discussed in **Chapter 10: In-Depth Interviews**, the questions you pose to participants—regardless of whether they are closed-ended or open-ended—should not lead subjects to answer in a certain way. Carefully scrutinize the phrasing of your questions to avoid biased language or a biased tone. In a similar way, try to minimize the inevitable problem of *social desirability bias*, the tendency for people to answer questions in a way that presents them in a favorable light. We all probably know the politically correct response to a variety of questions, whether or not we agree with the politically correct response. But we can phrase our questions in a way that does not make the "right" answer so obvious. As we discussed earlier in regards to in-depth interviewing, one way to encourage respondents who might be reluctant to give a stigmatized answer is to preface the question with a statement that suggests multiple responses are socially acceptable—for instance, "Some people think [*alternative 1*] . . . Other people think [*alternative 2*] . . . What do you think?" Reminding respondents of the study's promises of anonymity or confidentiality can also help assure respondents that they can speak openly and truthfully. Another way to avoid problems of social desirability bias is to try to phrase difficult questions in the most benign way possible. Earl Babbie (2016) recommends imagining how you would feel responding to your survey questions. If you would be uncomfortable, chances are others would as well.

1. Although this is generally true, some researchers argue that negatively worded questions should be integrated with positively worded questions to ensure that respondents have actually read each question with care. See, for example, this handout by Utah State University researchers on how to write survey questions for program evaluations.

Note that the leading nature of certain questions can be subtle. For instance, the Pew Research Center randomized the order of its questions on a survey that asked people whether they support either (1) maintaining spending in a specific area or (2) cutting that spending to reduce the budget deficit (Tyson and Doherty 2013). As summarized in **Figure 13.4**, the survey asked people about three spending areas—Social Security and Medicare benefits, military spending, and spending for the poor and needy—and presented the choice about whether to reduce or maintain spending in each of three areas in a random order. Pew found that people's support for cutting spending in a particular area was highest when that question was asked first out of the three questions; it decreased, though, when the question about spending reductions in that area was asked later. For instance, when respondents were asked whether to cut Social Security and Medicare, 63 percent felt current spending levels should be maintained when that spending area was mentioned first; 72 percent felt that way when it was mentioned second or third, however. It seemed that people were more willing to slash a particular set of programs at first blush, but after they heard about competing priorities, they became more reluctant to do so. The way that question ordering can influence how people respond on surveys sometimes leads survey researchers to randomize that ordering, which can be applied to a questionnaire's sections, to questions within a section, or even to the response options for a single question.



Question Order Effects in Deficit Polling

The three questions below were read to respondents in a randomized order ...

... Responses to the three questions differed depending on whether respondents heard a question 1st, or 2nd or 3rd

	Overall results	Results among those who heard question ... 1st	2nd or 3rd
What is more important ...	%	%	%
Taking steps to reduce the budget deficit	23	30	19
Keeping Social Security and Medicare benefits as they are	69	63	72
Taking steps to reduce the budget deficit	51	55	49
Keeping military spending at current levels	40	35	43
Taking steps to reduce the budget deficit	33	36	31
Keeping spending for poor and needy at current levels	59	57	60

Survey conducted Dec. 3-8, 2013.

PEW RESEARCH CENTER

Figure 13.4. Question Order Effects. What question precedes another in a poll will influence how people respond, as illustrated in this study by the Pew Research Center (Tyson and Doherty 2013). For instance, people's support for cutting spending in Social Security and Medicare was highest when they were asked first about that spending area, but it decreased if they were asked about cuts in other areas first. *Pew Research Center*

Response Options

While posing clear and understandable questions in your survey is certainly important, so, too, is providing respondents with unambiguous *response options*, the answers that you provide to the people taking your survey. In a quantitative survey, most, if not all, of your questions will be closed-ended in this fashion. This allows you to easily analyze your respondents' answers to individual questionnaire items and identify patterns in those responses. As we discussed in **Chapter 7: Measuring the Social World**, you can choose from many well-vetted types of closed-ended survey questions, such as Likert scales and semantic differentials.

Usually, you will want to ask respondents to choose a single (or best) response to each question you pose. That said, in some cases it makes sense to instruct respondents to choose two or more response options for a single question—for instance, if they are describing their racial and ethnic identity. One word of caution is that accepting multiple response options for a single question will make things more complex when it comes to tallying and analyzing your survey results—in many data analysis programs, you will need to treat every response option as a separate yes/no question.

There are two major guidelines to follow when devising response options for closed-ended survey questions. First, be sure that your response options are **mutually exclusive**. Look at **Figure 13.4**, which contains a GSS question asking about the respondent's family income. Notice how there are no overlapping categories in the response options. In fact, the dollar amount for each income category ends in "999" so that the next income category can start with the round number, ensuring that a single respondent can never fall into two categories. Meanwhile, the first category ("Under $1,000") and the last category ("$170,000 or over") cover all possibilities lower or higher than the income range captured by the other response options.

INCOME16: Categorical (Single)

In which of these groups did your total family income, from all sources, fall last year -- 2021 -- before taxes, that is.

Your total income includes interest or dividends, rent, Social Security, other pensions, alimony or child support, unemployment compensation, public aid (welfare), armed forces or veteran's allotment.

Categories:

{a}	Under $1,000
{b}	$1,000 to 2,999
{c}	$3,000 to 3,999
{d}	$4,000 to 4,999
{e}	$5,000 to 5,999
{f}	$6,000 to 6,999
{g}	$7,000 to 7,999
{h}	$8,000 to 9,999
{i}	$10,000 to 12,499
{j}	$12,500 to 14,999
{k}	$15,000 to 17,499
{l}	$17,500 to 19,999
{m}	$20,000 to 22,499
{n}	$22,500 to 24,999
{o}	$25,000 to 29,999
{p}	$30,000 to 34,999
{q}	$35,000 to 39,999
{r}	$40,000 to 49,999
{s}	$50,000 to 59,999
{t}	$60,000 to 74,999
{u}	$75,000 to $89,999
{v}	$90,000 to $109,999
{w}	$110,000 to $129,999
{x}	$130,000 to $149,999
{y}	$150,000 to $169,999
{z}	$170,000 or over

Figure 13.5. Mutually Exclusive Response Options. In this question from the 2022 GSS questionnaire about family income, the response options cover all possibilities—from no or negative income (captured by the "Under $1,000" option) to any income beyond the top $170,000 figure ("$170,000 or over"). *NORC*

Second, the response options for closed-ended questions should also be *exhaustive*. In other words, every possible response should be covered in the set of response options that you provide. (Sometimes, this criterion is called **collectively exhaustive**, leading to the useful mnemonic that survey questions must be MECE, or "mee-cee": mutually exclusive and collectively exhaustive.) In **Figure 13.6**, the response options for Questions 2 and 8 clearly do not cover all possibilities. Conversely, Question 4 has sufficient coverage across its response categories: those vehemently opposed to the statement can choose the first response option ("strongly disagree"), whereas those vehemently in favor can

choose the fourth response ("strongly agree"). All the possibilities in between these two extremes are covered by the middle two response options. An easy way of making a question's response options exhaustive is to include an "Other" category that captures any answers not covered by the explicit response options. That said, your stated options should offer sufficiently comprehensive coverage so that not too many people choose the "Other" option—as might happen with Question 1 in the example.

NEWS

‘The Onion’ Is Withholding Our Endorsement For President Until Both Candidates Respond To Our Questionnaire

Name (please print):

1. What do you consider to be the most important issue facing our nation?

1. Coal prices
2. Other

2. How can the debt of the U.S. government be eliminated?

1. Cut funding for public education after grade 2
2. Purge all federal monuments of salable marble
3. Incentivize taxpayers with immediate implementation of the Taxation Or Sterilization Choice Act
4. Retake the senselessly ceded Panama Canal Zone

4. Foreigners living within the borders of the United States, and anywhere else in the world, must be made to learn and speak English.

Strongly disagree / Disagree / Neither agree nor disagree / Agree / Strongly agree

8. What is the place of a woman in American society?

1. The home

Figure 13.6. Collectively Exhaustive Response Options. These are some questions from an *Onion* questionnaire for U.S. presidential candidates in the 2016 election. For Questions 2 and 8, the response options are not collectively exhaustive, as they don't cover all the answers that respondents would give. Thanks to the use of an "Other" category, Question 1 is technically exhaustive, but it still seems to be lacking something. *The Onion*

Sometimes survey researchers include **open-ended questions** in their questionnaires to gather additional details from respondents. An open-ended question does not include response options; instead, respondents are asked to reply to the question in their own way, using their own words. These questions are generally used to go beyond the closed-ended questions and to find out more about a particular respondent's experiences or feelings. If, for example, a survey includes closed-ended questions asking respondents to report on their involvement in extracurricular activities during college, an open-ended question could ask respondents why they participated in those activities or what they gained from their participation. While responses to such questions may also be captured using a closed-ended format, allowing participants to share some of their responses in their own words can make the experience of completing the survey more satisfying. It can also reveal new motivations or explanations that had not occurred to the researcher.

In the previous section, we discussed double-barreled questions. Response options can also be double-barreled, and this should be avoided. **Figure 13.7** is an example of a question that uses double-barreled response options.



Figure 13.7. Double-Barreled Response Options. Avoid phrasing your questions or your response options in a "double-barreled" fashion: when you ask respondents about two attributes at the same time, as in this example. This prevents your respondent from giving a mixed response—such as believing that college classes were "more demanding" but "less interesting."

Fence-sitting and *floating* can also present problems when you are crafting response options for your survey questions. **Fence-sitters** are respondents who choose neutral response options instead of disclosing their real preferences. This can occur if respondents are given, say, five rank-ordered response options, such as strongly agree, agree, no opinion, disagree, and strongly disagree (a Likert scale with five options). Some people will be drawn to the neutral option—"no opinion"—even if they actually do have an opinion. This is particularly likely to be the case if their true opinion is not a socially desirable opinion.

Floaters are respondents who choose a substantive answer to a question when, in actuality, they don't understand the question or don't have any opinion on the matter. Floaters can be motivated by the social desirability of having an opinion on an important issue—perhaps encouraged to do so by an enthusiastic interviewer they want to impress. This tendency can be exacerbated if a given question does not provide a neutral response. For instance, if a respondent is only given four rank-ordered response options, such as strongly agree, agree, disagree, and strongly disagree (a Likert scale with four options), those who have no opinion have no choice but to select a response that suggests they do have an opinion.

As you can see, floating is the flip side of fence-sitting, and there is no way to solve one problem without worsening the other. Ultimately, which and how many response options you provide to respondents depend on the goals of your research. Sometimes, researchers actually want to learn something about people who claim to have no opinion. In this case, allowing for fence-sitting would be necessary. If researchers are confident that their respondents are familiar with every topic in their survey, however, they may want to force respondents to choose an opinion.

A final bit of advice about thinking up response categories is to consider using a matrix to streamline the options you offer, especially for self-administered surveys. A **matrix** presents a set of questions for which the answer categories are all the same. As you can see in the example matrix in **Figure 13.8**, you might have a list of statements in your matrix and give the respondent the option of indicating that they strongly agree, agree, disagree, or strongly disagree with each of them. If you have a set of questions for which the response options are more or less the same, it may make sense to create a matrix rather than posing each question and its response options individually. You might need to revise the wording of some questions so that all of them can be answered with the same set of response options, but this approach has a substantial payoff. Not only will it save you some space in your survey, but it will also help respondents breeze through your questions.

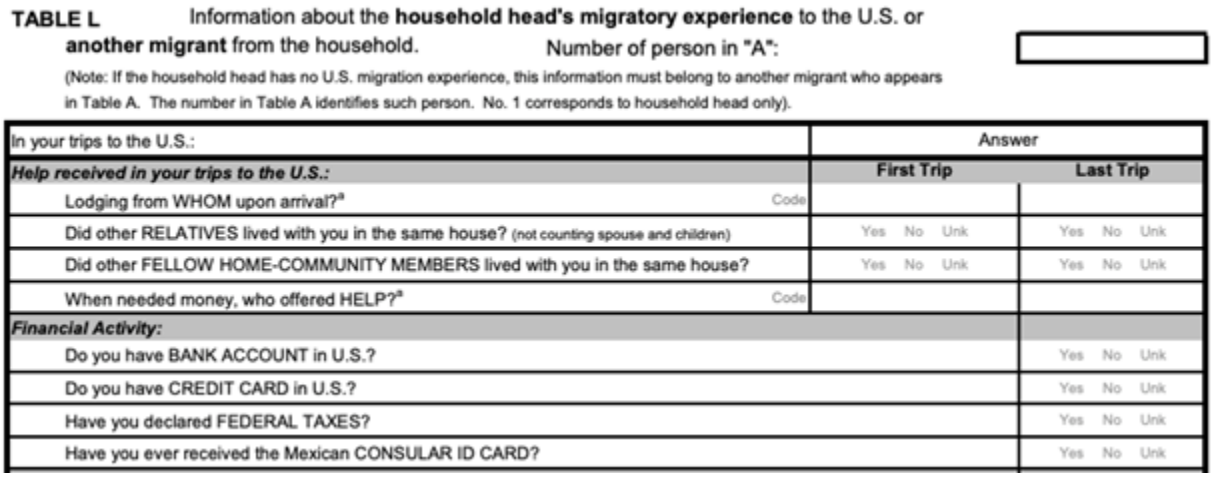


TABLE L Information about the **household head's migratory experience** to the U.S. or **another migrant** from the household. Number of person in "A":

(Note: If the household head has no U.S. migration experience, this information must belong to another migrant who appears in Table A. The number in Table A identifies such person. No. 1 corresponds to household head only).

In your trips to the U.S.:	Answer	
Help received in your trips to the U.S.:	**First Trip**	**Last Trip**
Lodging from WHOM upon arrival?[a] Code		
Did other RELATIVES lived with you in the same house? (not counting spouse and children)	Yes No Unk	Yes No Unk
Did other FELLOW HOME-COMMUNITY MEMBERS lived with you in the same house?	Yes No Unk	Yes No Unk
When needed money, who offered HELP?[a] Code		
Financial Activity:		
Do you have BANK ACCOUNT in U.S.?		Yes No Unk
Do you have CREDIT CARD in U.S.?		Yes No Unk
Have you declared FEDERAL TAXES?		Yes No Unk
Have you ever received the Mexican CONSULAR ID CARD?		Yes No Unk

Figure 13.8. Survey Questions in a Matrix Format. To streamline your survey questionnaire, you can group together questions with the same response options within a *matrix*. The excerpt presented here is from the Mexican Migration Project's ethnosurvey. Questions regarding migration-related help and financial activity that all had the same response categories (*yes*, *no*, and *unknown*) were placed within the same table. Matrices are also used frequently to list several statements and then allow respondents to agree or disagree with each statement using a shared Likert scale. *Mexican Migration Project*

Designing Questionnaires

In addition to devising good questions and response options, you'll also need to think about how to present them to survey respondents. In general, your questions should flow logically from one to the next. One of the first things to do once you've come up with a set of survey questions you feel confident about is to group those questions thematically. If you organize your questionnaire by topics and move similar questions next to each other, it will help the respondent understand where the survey is headed at any point in the process. It will also become readily apparent to you which of your questions repeat or overlap one another and therefore could be consolidated or dropped.

Once you have grouped similar questions together, you'll need to think about the order in which to present those

question groups. Many survey researchers believe that it is best to begin a survey with questions that will make respondents want to keep plowing through the questionnaire (Babbie 2016; Neuman 2009). In other words, don't bore respondents, but don't scare them away either.

To achieve the best response rates, some survey methodologists recommend that the questions should flow from the least sensitive to the most sensitive, from the factual and behavioral to the attitudinal, and from the more general to the more specific (Dillman, Smyth, and Christian 2014). In the list that follows, we offer additional advice, but note that the specific ordering may not matter all that much. You should really follow whatever sequence you think works well for the population who will be filling out your questionnaires.

1. **Start with nonthreatening questions that can be easily answered.** Basic demographic questions (e.g., age, gender, education level) are a good place to begin. If your survey deals with a sensitive or difficult topic, you don't want to scare respondents away or shock them by jumping right into your most intrusive questions.
2. **Never start with an open-ended question.** On a survey, it is generally better to use an open-ended question when you want to give the respondent a chance to elaborate on an answer they gave to a closed-ended question.
3. **When appropriate, follow a chronological ordering.** For instance, if your questions are focused on a sequence of events, it makes sense to start with questions about the earliest events and end with questions about the latest events.
4. **Ask about one topic at a time.** When switching topics, use a transition, such as "The next section examines your opinions about . . ."
5. **Use filter questions to avoid asking people about topics irrelevant to them.** If you have a series of questions that apply to just a portion of your respondents, you should include **filter questions** (also called *contingency questions*) in your questionnaire right before that set of questions. As illustrated in **Figure 13.9**, filter questions identify some subset of survey respondents who are then asked additional questions not worth asking of the entire sample. For instance, your filter question might ask if the respondent is a parent. If the respondent answered "no," the interviewer would skip over a subsequent set of questions asking about parenting. (In a self-administered questionnaire, you might provide instructions like the following: "If you answered 'yes' to question 5, please proceed to Section 2. If you answered 'no,' go to Section 3.") Routing respondents away from unnecessary questions will shorten the time it takes them to fill out the questionnaire and therefore reduce any survey fatigue they may be feeling.

Section R: Work and Pensions

INTERVIEWER CHECKPOINT: WHO IS PROVIDING INFORMATION (RESPONDENT/SPOUSE/PARTNER)?

Working now? (ask for respondent and spouse/partner)
- No:
  - Any work for pay?
  - Unemployed and looking for work last year?
  - Skip to "Employment History, Not Working Now"
- Yes:
  - Unemployed and looking for work last year?
  - Work for self/someone else/other?
    - Self-employed: Work connected with a personal business not reported earlier?
      - Yes: value of business, share, tax basis,
  - Industry and occupation
  - Normal hours/weeks worked
  - Pay (wages and other compensation)
  - Offered stock options?
  - Number of workers in business
  - Years in job
  - Years expect to continue in job
  - Covered by union/professional association contract?
  - Have disability insurance?
  - Included in any pensions?
    - No: Employer offers plans?/Eligible to be included?/Types offered
    - Yes:
      - How many different plans? (up to 2)
      - Currently receiving any retirement payments?
      - Plans not receiving benefits from?

Figure 13.9. Filter Questions. An outline for a set of questions posed in the 2016 Survey of Consumer Finances, a national survey sponsored by the U.S. Federal Reserve Board that asks respondents about their income, assets, and other aspects of their household finances. Note that this section of the questionnaire routes people to different questions based on how they respond to a filter question about whether they are currently working. *U.S. Federal Reserve*

How long should your questionnaire be? A perfectly good survey can be just a page or two, or it could be a dozen or more pages. The ideal length depends on several factors. First, how complex is your topic? Wanting to understand how grades

vary by gender and year in school certainly requires fewer questions than wanting to know how people's experiences in college are shaped by demographic characteristics, college attended, housing situation, family background, college major, friendship networks, and extracurricular activities. Even if your research question requires that many questions be included in your questionnaire, do your best to keep the questionnaire as brief as possible. Any hint that you've needlessly included questions will turn off respondents and may make them not want to complete your survey.

Second, how much time are your respondents likely to be willing to spend completing your questionnaire? If you are surveying retired persons, they may have plenty of unstructured free time to devote to a lengthy survey, but the same is probably not true for busy professionals. If they even agree to use their precious free time to complete a questionnaire, they may not want to spend more than a few minutes on it. If, however, you catch them at a work conference where they are milling about aimlessly anyway, maybe they'll give your survey more time. Likewise, if you are studying students and have official approval to administer your survey in class, you may be lucky to obtain a dedicated block of time just for your research.

As you can see, the time that researchers ask respondents to spend filling out surveys varies greatly. As with question order, there is no clear-cut, always-correct answer about questionnaire length. You'll need to consider the unique characteristics of your study and your sample and plan accordingly.

You can roughly gauge the amount of time it takes to complete your questionnaire through pretesting. **Pretesting**, also known as *piloting*, is a trial run of the questionnaire. It allows the researcher to get feedback so that the survey instrument can be improved before it is actually fielded. Pretesting can be quite expensive and time-consuming if your goal is to test your questionnaire on many people who closely resemble your target population. But you can learn a lot and spot potentially disastrous problems with your draft questionnaire by pretesting with a small number of people you have easy access to, even if they don't perfectly match your target population. (You can even ask friends and colleagues to complete the survey as though they were actually members of your population of interest.) Any helpful pretester can give you feedback on how understandable your questions are, whether the ordering of your topics flows smoothly, and whether any of your questions are exceptionally boring or offensive. Just by running through the questionnaire, they can also tell you where you should include filter questions or adjust those you already have. You should time pretesters as they take your survey. This will let you know whether you have some wiggle room to add additional items, or whether you really do need to trim your draft questionnaire. It will also give you a time estimate to provide respondents when you actually start administering your survey in the field.

A final consideration has to do with the design of self-administered questionnaires. If you are giving a respondent a printed questionnaire form or directing them to an online form, you need to ensure that it is exceedingly easy to read and navigate. A messy or unattractive presentation style can confuse or annoy respondents and make them take your survey less seriously. Avoid cramming too much material into a single page. Leave a reasonable amount of white space between items. Make sure all instructions are exceptionally clear. Proofread the form and fix any typos and bad grammar. Number your pages and include section headers to break up long lists of questions. And use a professional font with an easily readable type size—at least 12 point. You want the look and feel of your questionnaire to convey all the thought and effort you put into creating your survey.

Key Takeaways

1. Write questions that are clear and succinct, avoiding jargon and culturally specific terms, double

negatives, double-barreled questions (and answers), and leading and otherwise biased questions.
2. The response options for your questions should be mutually exclusive and exhaustive. Be aware, too, of how offering a neutral option might encourage or discourage people from offering their actual opinions (or lack thereof).
3. Make sure that your survey questions will be relevant to all the respondents answering, using filter questions when necessary to allow respondents to skip questions that don't apply to them.
4. Having people pretest your survey is essential. It will flag errors that may greatly reduce the value of your survey, and it will let you know whether respondents will be able to complete your survey in a reasonable amount of time.

Exercise

Based on the guidelines in this section, write 10 closed-ended questions for a survey on a topic of your choosing. Have a peer pretest your survey and give you feedback on what can be improved.

13.4. References

Babbie, Earl. 2016. *The Practice of Social Research*. 14th ed. Boston: Cengage Learning.

Belkin, Lisa. 2003. "The Opt-Out Revolution." *New York Times*, October 26.

Brockell, Gillian. 2020. "The History of Pollsters Blowing It Decades before Trump vs. Biden." *Washington Post*, November 6.

Cohn, Nate. 2022. "Who in the World Is Still Answering Pollsters' Phone Calls?" *New York Times*, October 12.

Curtin, R., S. Presser, and E. Singer. 2000. "The Effects of Response Rate Changes on the Index of Consumer Sentiment." *Public Opinion Quarterly* 64(4):413–28. doi: 10.1086/318638.

Dillman, Don A., Jolene D. Smyth, and Leah Melani Christian. 2014. *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. 4th ed. Hoboken, NJ: Wiley.

Fischer, Claude S., and Michael Hout. 2006. *Century of Difference: How America Changed in the Last One Hundred Years*. New York: Russell Sage.

Guo, Jeff, Nick Fountain, Emma Peaslee, and Jess Jiang. 2022. "Planet Money Tries Election Polling." *NPR*, November 4.

Harding, David J., Jeffrey D. Morenoff, and Jessica J. B. Wyse. 2019. *On the Outside: Prisoner Reentry and Reintegration*. Chicago: University of Chicago Press.

Kagay, Michael R. 1994. "Poll on Doubt of Holocaust Is Corrected." *New York Times*, July 8.

Keeter, Scott, Courtney Kennedy, Michael Dimock, Jonathan Best, and Peyton Craighill. 2006. "Gauging the Impact of Growing Nonresponse on Estimates from a National RDD Telephone Survey." *Public Opinion Quarterly* 70(5):759–79.

Kennedy, Courtney, and Hannah Hartig. 2019. "Response Rates in Telephone Surveys Have Resumed Their Decline." *Pew Research Center*. Retrieved October 11, 2023 (https://www.pewresearch.org/short-reads/2019/02/27/response-rates-in-telephone-surveys-have-resumed-their-decline).

Lepore, Jill. 2015. "Are Polls Ruining Democracy?" *New Yorker*, November 8.

Meyer, Vincent Maurice, Stan Benjamens, Mostafa El Moumni, Johan F. M. Lange, and Robert A. Pol. 2022. "Global Overview of Response Rates in Patient and Health Care Professional Surveys in Surgery." *Annals of Surgery* 275(1):e75–81. doi: 10.1097/SLA.0000000000004078.

Mortimer, Jeylan T. 2005. *Working and Growing Up in America*: Cambridge, MA: Harvard University Press.

Neuman, W. Lawrence. 2009. *Social Research Methods: Qualitative and Quantitative Approaches*. 7th ed. Essex, UK: Pearson.

Pager, Devah. 2003. "The Mark of a Criminal Record." *American Journal of Sociology* 108(5):937–75. doi: 10.1086/374403.

Percheski, Christine. 2008. "Opting Out? Cohort Differences in Professional Women's Employment Rates from 1960 to 2005." *American Sociological Review* 73(3):497–517. doi: 10.1177/000312240807300307.

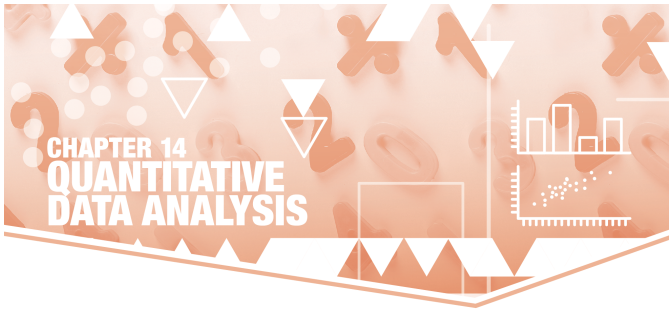
Pew Research Center. 2012. "Assessing the Representativeness of Public Opinion Surveys." *Pew Research Center*. Retrieved October 11, 2023 (https://www.pewresearch.org/politics/2012/05/15/assessing-the-representativeness-of-public-opinion-surveys).

Saad, Lydia. 2023. “More in U.S. See Unions Strengthening and Want It That Way.” *Gallup.com*. Retrieved October 13, 2023 (https://news.gallup.com/poll/510281/unions-strengthening.aspx).

Sullivan, Margaret. 2020. “We Still Don’t Know Much about This Election—except That the Media and Pollsters Blew It Again.” *Washington Post*, November 4.

Tyson, Alec, and Carroll Doherty. 2013. “Polling on the Deficit: Why Question Order Matters.” *Pew Research Center*. Retrieved October 16, 2023 (https://www.pewresearch.org/short-reads/2013/12/20/polling-on-the-deficit-why-question-order-matters).

14. QUANTITATIVE DATA ANALYSIS



Numbers pervade our lives. For one thing, we're bombarded with them in the mass media, as seen in the following examples from news articles:

- "Incidents of white supremacist propaganda distributed across the nation jumped by more than 120% between 2018 and last year, according to the Anti-Defamation League, making 2019 the second straight year that the circulation of propaganda material has more than doubled" (Schor 2020).
- "More than 20% of black respondents were in families with [worries about their medical bills]. That compares with 13% of whites, more than 15% of those who identified as Hispanic and 7% of Asians" (Murphy 2020).
- "In the new study, among the 84 [Covid] patients who took hydroxychloroquine, 20.2% were admitted to the ICU or died within seven days of taking the drug. Among the 97 patients who did not take the drug, 22.1% went to the ICU or died. The difference was not determined to be statistically different" (Nigam and Cohen 2020).

Sociologists publish a lot of research involving quantitative data analysis. As an informed reader of this literature, you should appreciate the many ways that quantitative data can be used in sociology and learn how to interpret simple analyses. At this point, however, we don't expect you to understand much of what is going on in the quantitative research articles you read. We also recognize how intimidating this work can be. Many students get anxious whenever math comes up, and they do their best to avoid courses that involve quantitative data analysis.

Luckily, if you're in that group, you can relax. Our goal for this chapter is to make you sufficiently familiar with statistical analysis so that you get some useful information from quantitative work—even from more complicated analyses. You won't need to do any calculations to understand this material. And there will be no formulas, we promise!

How do sociologists manage to do all of the mathematical calculations needed to process, say, a survey that involves data collection from hundreds or even thousands of people? The answer is that they don't—data analysis programs do that work for them. This is another reason that you should not be afraid of quantitative data analysis even if you are math-phobic. For basic applications, the software handles most, if not all, of the numerical heavy lifting, meaning you often just need an intuitive understanding of quantitative procedures to interpret or even generate useful results.



Your (much older) professors may have learned to be less afraid of math by watching educational programs like *Schoolhouse Rock!* Although we (thankfully) don't have any catchy song-and-dance routines for you, we hope to bring the same calm and reassuring approach to our review of statistical analysis in this chapter. Don't fixate on the math. Instead, try to understand intuitively what researchers are trying to do with each data analysis technique. *Scholastic Rock, Inc./ABC, via Schoolhouse Rock Fandom*

For the rest of this chapter, we're going to use SPSS, the Statistical Package for the Social Sciences, to walk you through quantitative data analysis. First marketed in 1968, SPSS is often available free to students in colleges and universities to use in their courses and research projects. The following chapter sections will refer to SPSS specifically, and the data will be processed in SPSS, but the discussion will apply to social scientific data analysis programs in general.

We'll start off by talking about how you should go about either entering the data from your own survey into SPSS or finding one of thousands of existing data files available for free on the internet. Once you have data to analyze, you should generate **univariate statistics** (statistics examining one variable at a time) to describe your data. We'll cover the logic behind this analysis and what the results should look like in the tables and charts you generate in SPSS. Next we'll talk about **bivariate statistics** (statistics examining the relationship between two variables). We'll go over a simple kind of bivariate analysis, crosstabulation. We'll discuss the logic of **inferential statistics**—determining whether a relationship between two or more variables truly exists—and apply the chi-square test of statistical significance to our crosstabulation results. We'll wrap up the chapter by briefly describing more advanced approaches to quantitative analysis. The goal in this chapter is to give you an intuitive understanding of when and how to apply these different statistical procedures, rather than trying to explain the math or statistical theory behind each procedure. (For that background, consult a statistics textbook.)

Statistical Analysis Programs

We find SPSS to be the easiest statistical analysis program for students to learn. It has a graphical user interface, so you don't have to type in commands as you do with most other data analysis programs. Note, however, that social scientists use a wide range of programs, so you will inevitably come across data files in other (non-SPSS) formats, charts that look different from the standard in SPSS, and other sorts of differences that might throw you off. That said, we believe if you understand one data analysis program well, you'll be able to at least grasp how other programs work (and remember from **Chapter 2: Using Sociology in Everyday Life** that employers see quantitative analysis skills as highly attractive in job candidates).

There are pros and cons to every statistical analysis program on the market. If you go the route of specializing in quantitative analysis, you should pick the one that suits you best. SPSS is commonly used in psychology, and Stata in economics. SAS is a standard at many government agencies and in the field of public health. (No particular program seems to dominate within sociology overall, though certain subfields have their favorites.) Colleges and universities often provide licenses for different data analysis programs to their students for free, just as employers do for their workers. The software licenses are very expensive and are not something that most students can or should pay for on their own.

In fact, if you are bothering to learn a different statistical analysis program than the easy-to-use SPSS, we'd highly suggest you learn R. This analysis platform, which is completely free and open-source, was created by University of Auckland researchers in 1993 and has exploded in popularity since, with a vibrant developer community that creates and freely shares modules to expand the functionality of the software. Quantitative sociologists and other researchers have adopted it in large numbers, using it for tasks ranging from basic statistical calculations to highly specialized analyses that use tailored modules. You can't go wrong learning R, particularly given its future prospects as the go-to data analysis program. However, R (at least in its plain-vanilla form) has a steeper learning curve than SPSS, so we use the latter program in this textbook (remember, we're trying really hard not to scare you!).

Opening chapter image credit: *Karolina Grabowska, via Pexels. Adapted by Bizhan Khodabandeh.*

14.1. Preparing for Data Analysis

Learning Objectives

1. Describe the data-entry process and the role of a codebook in guiding that process.
2. Explain how you can find existing datasets and the reasons you might conduct a secondary data analysis.
3. Describe the format of a dataset in SPSS (or whichever analysis program you are using).

Before you can begin any quantitative analysis, you need to place the data you wish to analyze into a **dataset**—a structured collection of data that is typically stored in one or more computer files (called *data files*). If you administered your own survey, you'll need to enter the information from your survey forms into a new data file. If you're using data that other researchers have collected (what we call **secondary data**), you will need to download the relevant data file or files along with any documentation that has been provided. In the following sections, we'll review these two approaches of creating a new dataset and using secondary data, respectively.

Creating a New Dataset

Whenever you administer a survey, it can be very exciting to receive those first few completed forms back from respondents. As the responses pile up, however, your emotions may shift from euphoria to dread. A mountain of data can be overwhelming. How can you pull a set of intelligible findings out of this stack of submitted forms?

One major advantage of quantitative methods such as survey research is that they enable researchers to describe large amounts of data, which can be represented by and reduced to numbers within a dataset. In order to condense your completed surveys into analyzable numbers, you'll first need to create a **codebook**. A codebook is a document that describes the content, structure, and layout of your dataset and outlines how the researcher translated the data from words into numbers—a process we call **coding**. Note that the "coding" we're doing here is different from the coding we do in qualitative data analysis. Even though both kinds of coding help us analyze our data, in this case, we are not focused on deciphering and interpreting meanings from text or images. Coding of quantitative data tends to be more formulaic—a matter of converting people's responses to survey questions (all or most of which were closed-ended to begin with) into numbers. The codebook records what specific response each of our numerical **codes** refers to.

Codebooks can range from relatively simple to quite complex. Let's discuss a fairly straightforward example of a codebook from the Mexican Migration Project, which has gathered data on the characteristics of migrants traveling from Mexico to the United States. As you can see in the codebook excerpt in **Figure 14.1**, the questions on the original survey (the survey **items**) have been converted into unique, single-word **variable names**. (As we'll discuss later, the variable name comes in handy when entering data into a computer program for analysis.) In turn, the response options for each question (called **attributes** when they refer to different categories of a variable) have been converted into numerical values. For example, the two questions at the bottom—labeled with the variable names GREEN and CITIZEN—ask whether the respondent is a "green card" holder (i.e., a legally recognized permanent U.S. resident) and a U.S. citizen, respectively. For both questions, if the respondent answered "yes," their response in the dataset was coded as 1; if they answered "no," it was coded as 0. This codebook doesn't contain the actual text for these two questions, but many codebooks will list the exact question wording after every variable name. Note that sometimes a survey item will have

multiple variables attached to it—for instance, if the question allowed the respondent to choose multiple response options, which each need to be coded as a separate variable.

At this point, you may be wondering why you should bother assigning numbers to simple response options like "yes" and "no." One reason is that it makes entering data into your dataset easier. When it comes time to input a respondent's answer into a data analysis program, you will just need to type "0" or "1," rather than "yes" or "no" (or perhaps a much longer answer, depending on the specific response option). Coding responses in this way also reduces errors. For instance, if you typed in "nope" rather than "no," your dimwitted computer will treat "nope" as a separate answer from "no."

In this example, only two response options were given to respondents for each question: yes or no. When a variable only has two response options—yes/no, true/false, and so on—we call it a **dichotomous variable**. These two questions—captured in the two variables GREEN and CITIZEN—are a special kind of dichotomous variable: a **dummy variable**. For dummy variables, the code "0" indicates the absence of a condition and "1" indicates its presence. For the first GREEN question, then, "1" refers to green card holders, and "0" refers to everyone else (U.S. citizens, other visa holders, unauthorized migrants, and everyone else in the sample who does not have a green card) . For the second CITIZEN question, "1" refers to U.S. citizens, and "0" refers to everyone else—people who are not U.S. citizens. Researchers often prefer using this "0" and "1" coding scheme for simple dichotomous variables. Not only does it make certain kinds of advanced analysis easier, it also makes the interpretation of the variable very straightforward: "1" is always the presence of the condition listed in the variable name (for instance, a green card holder for GREEN, or a U.S. citizen for CITIZEN).

LIFE Data File: Variable Codes

Variable	Label	Value	Definition
USDOC	**U.S. Documentation**		
		0	N/A (Not in U.S. during person-year)
		1	Legal resident
		2	Contract - Bracero
		3	Contract - H2A (agricultural)
		4	Temporary: Worker
		5	Temporary: Tourist/visitor
		6	Citizen
		7	Silva Letter
		8	Undocumented (includes false documents)
		9	Refugee / asylum
		10	PTAT
		12	DACA
		8888	N/A (never migrated to the US)
		9999	Unknown
GREEN	**Green card?**		
		0	No
		1	Yes
		9999	Unknown
CITIZEN *Note: Question introduced in community #72*	**U.S. Citizen**		
		0	No
		1	Yes
		9999	Unknown
		.	Communities 1 – 71

Figure 14.1. Codebook Excerpt from the LIFE Dataset of the Mexican Migration Project. Coding is the process of converting data into a numeric format. Researchers create codebooks to guide the coding process. A codebook describes each variable in a dataset, including the *items* (questions) associated with that variable and the numerical *codes* assigned to each response category. *Mexican Migration Project*

Following the example in **Figure 14.1**, you can create your own codebook in a word-processing document or spreadsheet. First, give each of your questions a unique variable name. Statistical analysis programs generally require the variable name to be a single word, so you will want to use something like GREEN or CITIZEN. If you need to use longer words or phrases, connect them with an underscore or just use abbreviations, such as PARTY_ID (for the political party a respondent identifies with) or POLORIENT (for their political orientation). As we noted earlier, codebooks often contain the full text of each question, so you can add it underneath the variable name. After the question text, you should list the question's response options and the numerical codes you have assigned to each one.

Note that although your codebook contains your variable names, in an actual paper, you will usually not want to list specific variable names (e.g., GREEN or CITIZEN). Those variable names mean something only to the researcher who is analyzing the dataset, so for public consumption, you generally just want to use a word or short phrase to describe the variable in the text: "the variable for permanent residency status" for GREEN, for instance, or "the variable for citizenship status" for CITIZEN. In special cases, researchers will mention variable names in a paper, and when they do so, they may choose to type out the names in all-caps (e.g., EDUC for education level) or in italicized lowercase letters (e.g., *educ*). We'll follow the latter convention throughout the rest of the chapter so that you don't feel like we're shouting at you!

Your survey will likely be missing responses from some respondents on some questions. That's okay. These nonresponses—what we call *missing data*—are unavoidable in large datasets. Respondents may not answer certain questions that are ambiguously worded or cover sensitive topics. They may accidentally skip a question—or even decide to give up on the survey altogether. During data entry, these nonresponses to questions should be categorized as **missing values**. (Missing values can be distinguished from **valid values**, which are the data related to the actual response categories for your questions—in other words, the data you care about.) In your codebook, you can flag missing values by using a numeric value such as −1 or 99 that will not overlap with any valid values. For instance, if you are adding a question about a respondent's age to your codebook, you probably wouldn't want to use 99 as a missing value, since a person might very well have an age of 99. Instead, you'd want to use 999 or −1 since an age of 999 or −1 would be nonsensical in most contexts. (In the codebook excerpt presented in **Figure 14.1,** the value for missing data is 9999.)

Codebooks will frequently list two or more missing values for a question. That's because in a survey interview, a respondent may fail to respond in a particular way: they may refuse to answer a question, for instance, or may tell the interviewer, "I don't know." Rather treating these two nonanswers in the same way, you might want to create separate codes for each—say, giving "refuse to answer" a code of 98 and "don't know" a code of 99.

As you decide on how to code response categories, keep in mind the levels of measurement we discussed in **Chapter 7: Measuring the Social World**: *nominal*, *ordinal*, and *scale*. (These are illustrated in **Figure 14.2**, which appeared in that earlier chapter.) If you are using a scale-level variable, you may not need to provide codes for that variable in your codebook other than noting any missing values. For instance, if your scale-level variable was annual household income, and the respondent answered $60,000, the number 60000 could be stored as it is in your dataset, with no coding necessary. But what about for a nominal-level variable like the region of the country where a person lives? Which response category should be "1," which should be "2," and so on? Should any region be coded as "0"?

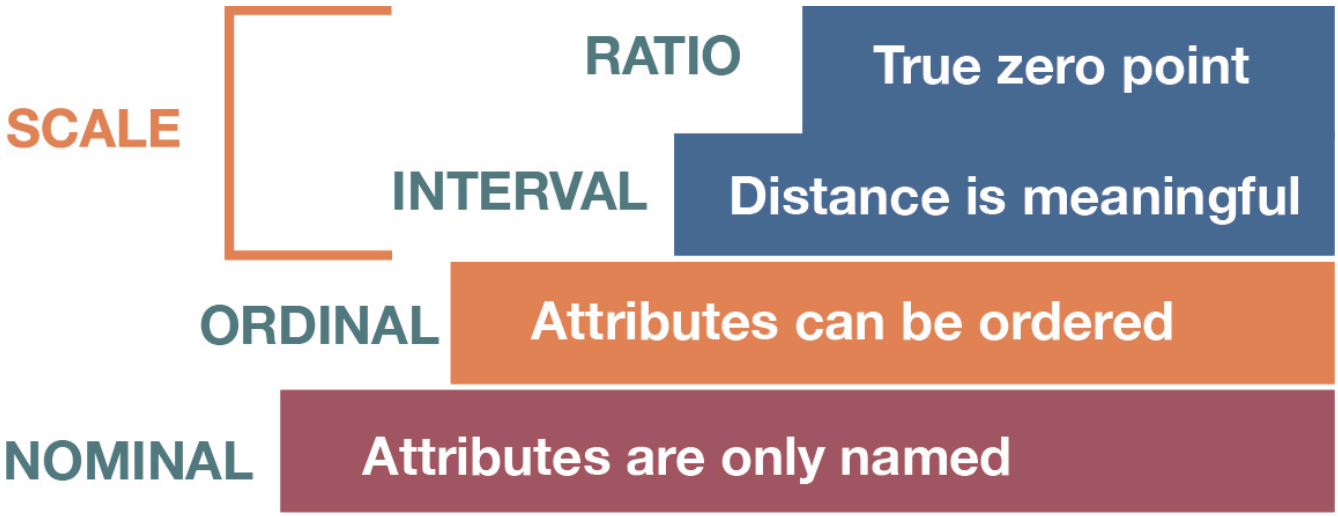


Figure 14.2. Levels of Measurement. The level of measurement for a variable will determine what statistical procedures we can apply to it during our analysis. As we described in Chapter 7: Measuring the Social World, the three overarching levels of measurement are *nominal*, *ordinal*, and *scale*. (See that earlier chapter for descriptions of these different levels of measurement and a discussion of *interval* and *ratio* levels of measurement, which both fall under the scale level of measurement.)

If you want to create a dummy variable out of your nominal-level variable, then, yes, you should use 1 for the region you want to highlight—say, the South—and 0 for everywhere else. Other than that specific yes/no scenario, though, the numbers you use don't really matter for nominal-level variables. **Figure 14.3a** shows one possible scheme that includes five different U.S. regions (five response categories). In a straightforward fashion, they are numbered in alphabetical order.

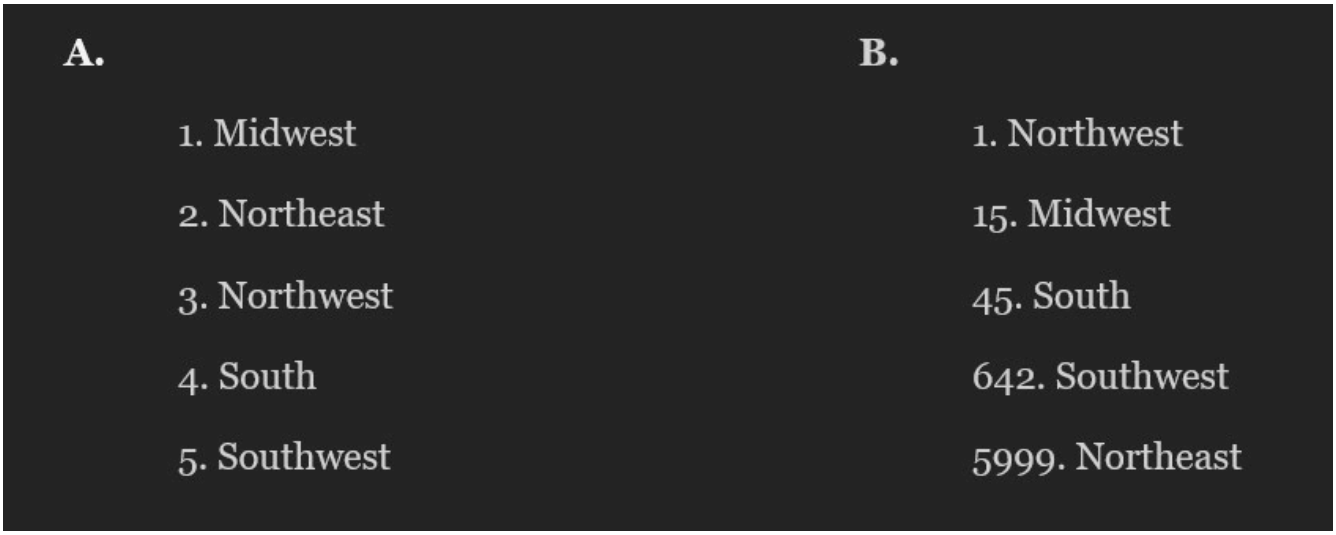


Figure 14.3. Numerical Codes for a Nominal-Level Variable's Response Options. Nominal-level variables don't have any inherent ordering to their response options. If your dataset contained a nominal-level variable about where in the United States the respondent lived, you could choose to order the response options alphabetically (or not), and you could choose to assign numerical codes from 1 on (or not). Following a logical ordering scheme will make it easier to understand your data, but it is not strictly required for nominal-level variables.

With a nominal-level variable like this, however, we could have gone with many other numbering schemes, including the one depicted in **Figure 14.3b**. For instance, we didn't really need to list the different response options in alphabetical order, given that there is no intrinsic ordering to nominal-level attributes (the response options). That said, the first numbering scheme clearly looks better, and for your own sanity, we advise arranging the response options in your codebook (and in your questionnaire) in a simple and logical fashion.

Next, let's look at ordinal-level variables. Consider a question that the General Social Survey (GSS) has asked in the past: "Do you believe there is a life after death?" This question could be answered in a variety of ways, and we might decide to simply provide our respondents with "yes" and "no" as response options. This question seems different, however, from the yes/no question we considered earlier that asked whether or not someone is a U.S. citizen. In this case, we might expect our respondents to have more or less intense beliefs about whether there really is life after death.



Most of the American adult population believes there is life after death. However, there are differing degrees of belief/disbelief that can be examined by using an ordinal-level variable. The GSS classifies the possible responses from high belief to high disbelief using four ordinal values: *yes, definitely*; *yes, probably*; *no, probably not*; and *no, definitely not*. *Alexander Grey, via Unsplash*

In their actual question, the GSS captures this intensity of belief by using response options that collectively represent an ordinal-level variable. The coding scheme for this variable is shown in **Figure 14.4a** (note that the last two codes are missing values). In this example, the valid values are ordered by intensity: the higher the number, the weaker the respondent's belief in life after death.

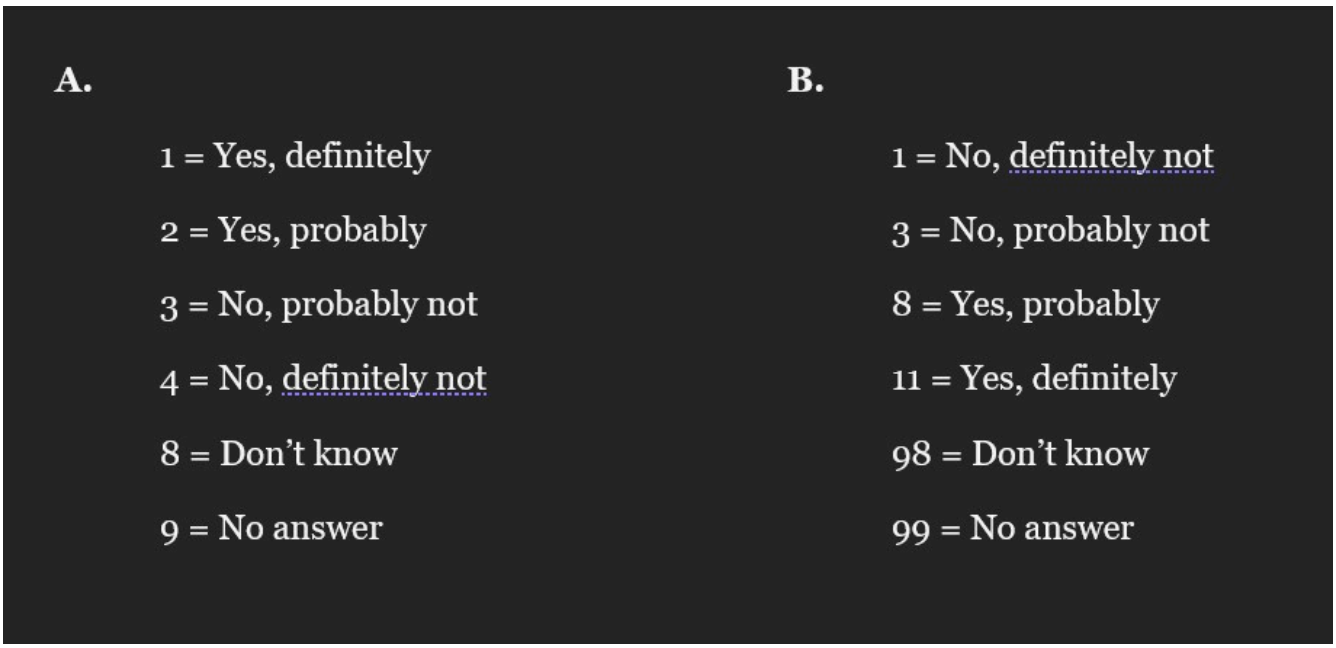


Figure 14.4. Numerical Codes for an Ordinal-Level Variable's Response Options. Unlike nominal-level variables, ordinal-level variables have a rank ordering—one response option comes before another, which comes before another, and so on. However, the distances between the attributes of an ordinal-level variable are not meaningful, so you could theoretically choose any numerical code for that item's response options, so long as it preserves the overall ordering from low to high or high to low. (Again, though, following a logical ordering scheme—as in Panel A—will make life easier for you!)

For ordinal-level variables, *the rank order of any valid values should be clear.* However, the GSS designers did not have to use the particular numbers in Panel A. They didn't have to have a gap of just one integer between every pair of numbers, and they could have made those valid values track in the opposite direction, with higher numbers representing *stronger* belief—as shown in **Figure 14.4b**. Of course, using numbers with different-sized gaps between them seems unnecessarily complicated, and ordinal-level variables are typically represented by consecutive integers in datasets. The point here is that the codes you choose for these categories are somewhat arbitrary.

Once you've created your codebook, what's next? If you've utilized an online tool such as SurveyMonkey to administer your survey, here's some good news—most online survey tools come with the capability of importing survey results directly into a data analysis program. If you've administered your questionnaire in paper form, however, you'll need to enter the data manually (see the sidebar **Deeper Dive: Entering Your Own Data in SPSS**).

For those forced to conduct manual data entry, there probably isn't much we can say about this task that will make you want to perform it other than pointing out the ultimate reward of having a dataset of your very own ready to analyze. At best, it is a Zen-like practice akin to raking sand. At worst, it is mind-numbingly boring. Although you can pay someone else to do your data entry for you (a common chore given to student research assistants), you should ask whether you trust that person to make no errors when entering your data. Even a few typos can jeopardize the results of your project. It may be worth your time and effort to do the data entry yourself.

Deeper Dive: Entering Your Own Data in SPSS

The following helpful series of videos explains how to enter your own data into SPSS. Total viewing time is approximately 17 minutes. We recommend following along in SPSS while you are watching **Videos 14.1–14.4**, pausing as needed to apply the suggested procedures to your own dataset.

Part 1 | Part 2 | Part 3 | Part 4

Secondary Data Analysis



Sociologists have access to many existing datasets through organizations like the Inter-university Consortium for Political and Social Research (ICPSR), which stores thousands of datasets on its servers and offers many of them for free to the public. *panumas nikhomkhai*, *via Pexels*

Secondary data analysis is analysis of data that has previously been collected and tabulated by other sources. In this age of **big data**, where a vast and growing assortment of information has been digitized, you can find and download datasets

on virtually any topic within minutes—many of them publicly accessible and free of charge. Government agencies at all levels routinely release data related to their operations—from city crime statistics, to monthly job reports from the U.S. Bureau of Labor Statistics, to country indicators of health and well-being compiled annually by the United Nations Development Program. In addition to the official data that governments release, you may be able to obtain data from companies and organizations that conduct surveys or release selected portions of the data they generate about users and clients. Individual researchers frequently make the data they have collected available to other scholars, too.

The main way that governments and other organizations provide their data is in an aggregated form—counts, averages, and other measures of anything from unemployment rates to the population's gender breakdown, reported in tables and charts. Although these figures can be useful, researchers often want to go beyond those aggregated results and look at the underlying data for themselves. They seek information on specific individuals (or whatever a dataset's unit of analysis is). This **microdata** allows them to conduct their own analysis of patterns within variables and relationships between them.

Many government agencies and other organizations provide microdata in datasets they post online or otherwise offer to researchers, and this is the type of data you should obtain to conduct the analyses we cover in this chapter. We should also note that governments and other organizations routinely release data based on their own surveys, but occasionally they also allow researchers to examine **administrative data**—data from actual official processes, such as registration records and customer transactions, that organizations use for their operations. In the last chapter, for instance, we discussed how a team of sociologists was able to convince Michigan authorities to let them analyze records relating to people released from prison (Harding, Morenoff, and Wyse 2019). The use of administrative data has revolutionized the study of topics like income inequality. That's because administrative data like tax records can provide much better estimates of people's actual earnings than surveys do, given the problems with social desirability and accurate recall that we've previously discussed.

If you are considering a quantitative research project, you should start by thinking about which datasets already exist on your topic. Have past surveys covered the concepts and relationships you want to examine? Have they sampled the target population you are most interested in? Would it be difficult for you to conduct your own survey that improves upon those prior efforts? If you are saying "yes" to all these questions, you will probably be better served using secondary data than doing your own data collection, which can be very costly in terms of time and money—especially when you need to generate a representative sample of large populations whose members are unknown to you.

Here are some of the key advantages of conducting secondary data analysis:

- **You don't have to spend time collecting the data.** Analyzing secondary data enables you to avoid the entire data collection process, which can significantly speed up your research. All that's required is that you have access to the data.
- **You can find datasets very easily online.** Government agencies typically share data with the public on their websites and in published form. Researchers will upload the datasets they analyzed for published work to online repositories or their professional websites. Frequently, a quick online search will locate relevant datasets for you to download. For example, a search for "adolescent risk behavior datasets" will lead you to the Youth Risk Behavior Surveillance System (YRBSS), a series of datasets compiled by the U.S. Centers for Disease Control and Prevention. Occasionally, getting access to a particular data file will require filling out online forms to identify yourself as a researcher and describe the study you are working on; less frequently, you may have to go through a formal process to apply for access (usually only for datasets with particularly sensitive personal information, such as health records). As we discuss later in the chapter, massive data archives also reside online, which you can search to locate datasets focused on very specific topics if a general web search turns up nothing. Finally, note that you can always ask an organization to share their data with you for research purposes. Sometimes, even corporations will share the microdata from surveys they conduct when the request has a scientific purpose.

- **You don't have to pay for most datasets.** When governments collect data, they usually make the data files available for free. And when they fund studies, the conditions of the grant often require researchers to do the same. For instance, the GSS, which we analyze later in this chapter, is funded by the National Science Foundation and can be downloaded for free. The majority of the thousands of datasets archived by the Inter-university Consortium for Political and Social Research (ICPSR) are free public-use files. The remainder can be downloaded by students, faculty, and staff at universities that have ICPSR memberships.
- **You can find data from rigorously conducted surveys whose samples would be hard or impossible to obtain on your own.** It is prohibitively expensive to survey large populations over wide geographic areas like a state or country. If you are conducting this research on your own, it is unlikely you'll have the resources to generate a representative sample. You may want to leave it to the pros. Highly experienced survey specialists oversee major ongoing projects like the GSS, using sophisticated techniques to ensure that their samples truly reflect the population of interest.
- **You will not have to bother with getting institutional review board approval.** Publicly available datasets are routinely anonymized, which means that researchers process the data to remove details that would allow individual respondents to be identified (see **Chapter 8: Ethics** for a discussion of the importance of protecting the identities of research participants). As a result, you usually do not need the institutional review board at your college or university to scrutinize the procedures of your study when you use secondary data (presumably, another institutional review board vetted the original survey). You should check with your institution to be sure.

As is true of all methodologies, secondary data analysis also has certain limitations, such as the following:

- **Organizations collecting secondary data may provide insufficient background information.** You will not be able to evaluate the quality of the sampling or other data collection procedures a survey engaged in unless the original researchers did a good job of explaining their methodology. Government agencies that specialize in data collection, such as the U.S. Census Bureau, offer extensive information on the design and implementation of their surveys. So do well-respected research outlets like the National Opinion Research Center, Gallup, and the Pew Research Center. But a random one-off study may or may not have had rigorous procedures in place—you can look through the accompanying documentation to get a sense of their methods, but you'll often be in the dark. Because administrative data is typically not meant for public consumption, it likely to come with very little documentation, which means you will need to make additional efforts to suss out the precise meaning of variables and response options or the organization's specific procedures for data collection.
- **You are limited to what questions have already been asked.** This is a key reason that researchers choose to collect their own data: the available secondary data does not match their research purposes. If you want to research the factors that might influence a young woman's decision to have a legal abortion, for instance, you would not want to look at the GSS. That survey contains many questions about people's attitudes about whether abortion should be legal, but it does not specifically ask how a respondent might approach that decision themselves. To pursue such a research question, you would need to find another survey that does a better job of measuring the relevant variables of interest—or you could field an original survey that asks the specific questions you want to ask.
- **The secondary data available to you doesn't measure what you want to measure in a valid way.** This limitation is similar to the last point, but worth mentioning separately. Say that you would like to assess people's attitudes about protecting the environment. You find a survey that does ask questions about the environment. However, it only asks respondents about climate change, recycling, and sustainable energy alternatives. These questions might be of interest to you, but unfortunately they would not provide you with a valid measure of a person's overall attitudes about protecting the environment, which should go well beyond those three areas.
- **Data may be in a different format than you require.** For example, you might want to study young adults between the ages of 18 and 25. However, you discover that the age variable in your secondary dataset is not measured in single years, but in ranges—and the first age category is 18 to 35. There is no way you could study your desired age range with that dataset.

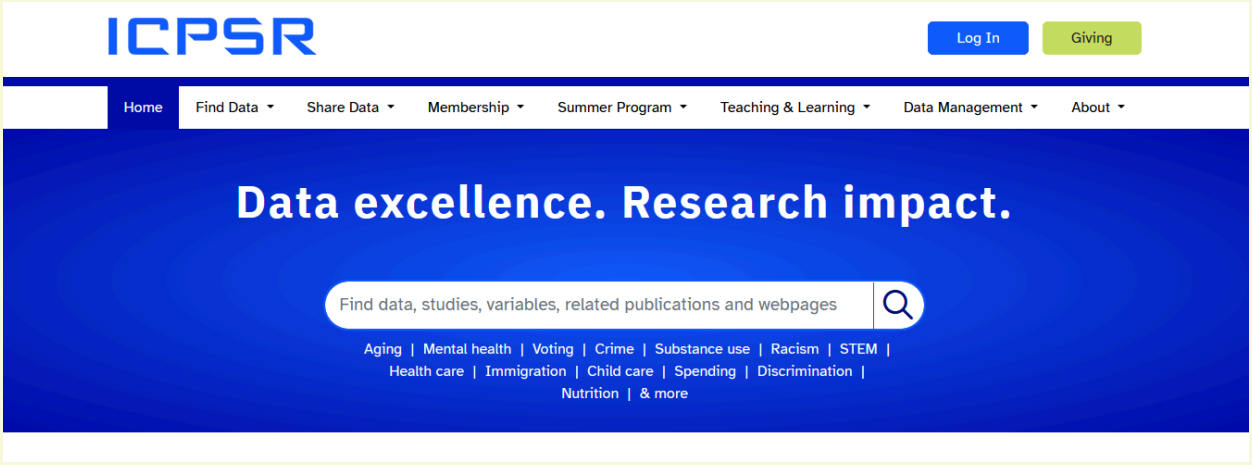
If you decide you want to use secondary data, you can pursue two different strategies to find relevant datasets. One is to find questions on your specific topic within a survey that asks respondents about a wide range of topics (what is called an **omnibus survey**). A good place to start is the list of popular datasets we have collected at this site. Some ask general questions, while others have a narrower focus—for instance, on health or finances. Regardless, this search strategy is about identifying a relatively small set of relevant questions within an expansive survey questionnaire.

Note that some long-running surveys will regularly or occasionally have modules or supplements on specific topics. For instance, the GSS asks many of the same questions every time it is fielded, but it will occasionally ask a set of questions about beliefs about science, views on the economy, or other topical issues just for that year's survey. Browse a survey's website to find out what modules or supplements it has done.

The second strategy for finding relevant secondary data is to search for a dataset focused more narrowly on your topic. In the data repositories section of our data page, you will find a number of online catalogs where you can search for and download relevant datasets. The sidebar **Searching a Data Repository** has instructions on how to use the repository maintained by ICPSR, which stores datasets on a vast array of social-science topics.

Regardless of which strategy you take to find your dataset, eventually you should be able to get your hands on both the survey's data files and its corresponding documentation. These should be available on the survey's website or its page within a data repository. Frequently, a survey's questionnaire is included or integrated within its codebook, though sometimes you will need to look for a separate questionnaire document in addition to the codebook (usually a PDF file). If it is available, you should download the version of the survey's data file that is in a format matching whatever statistical software package you are using to analyze it. (If you are using SPSS, for example, you should make sure your data file has an SAV file extension.) If you can't find the right format, it is possible to use the same data in another format. SPSS and other analysis packages can read data in a raw form—that is, comma-separated values (CSV) data files used for simple spreadsheets—and you can also convert your data from one software package's format to another format.

Searching a Data Repository



You can sign up for an account on the ICPSR website, which enables you to download datasets for free. ICPSR offers many of its datasets in SPSS format, but you'll need to check the specific dataset you want to analyze to verify what formats it is in. (Note that you can convert a data file from one data analysis program's format to another format, or you can download the data as a raw CVS data file.)

You will need to register for a free ICPSR account to access the data stored in this voluminous repository. If you are affiliated with a college or university, make sure to use your institutional email address so that you can obtain access to more of the data. During the registration process, you will be asked if you want to "allow your campus Official Representative (OR) to view your name and email address"; choose whichever option you prefer. Note that you will need to click on a link in the email that ICPSR sends you (which might go to your spam folder) to activate your account.

To find datasets, enter keywords into ICPSR's search engine (under **Find Data**). To learn more about a study, click on its name in the search results. You should skim the description of the study, particularly any details about the survey's sample and its questions. You also should look at the survey's questionnaire or codebook. Questionnaires and codebooks are usually provided on the study's ICPSR page, or you can just search online for them with the name of the study and "questionnaire" or "codebook" as keywords.

On ICPSR, you can choose to download the dataset along with the study's codebook all at once, or you can download the codebook by itself (which makes sense if the dataset is large). Click on the **Download** button and select SPSS (or whichever statistical analysis software you are using). If there is a list of multiple datasets and codebooks to choose from, pick the one that seems most relevant to you, choose the most recent year, or select whatever general option includes most or all of the data (e.g., an option called "main data," "main study," or "public-use data").

You will need to click **Agree** on the Terms of Use page that appears. Then the files will be downloaded in a compressed file. You will need to extract its contents. In Windows, just right-click the file and then select **Extract All**. On a Mac, double-click the file to extract it. The extracted data file will appear in a folder called

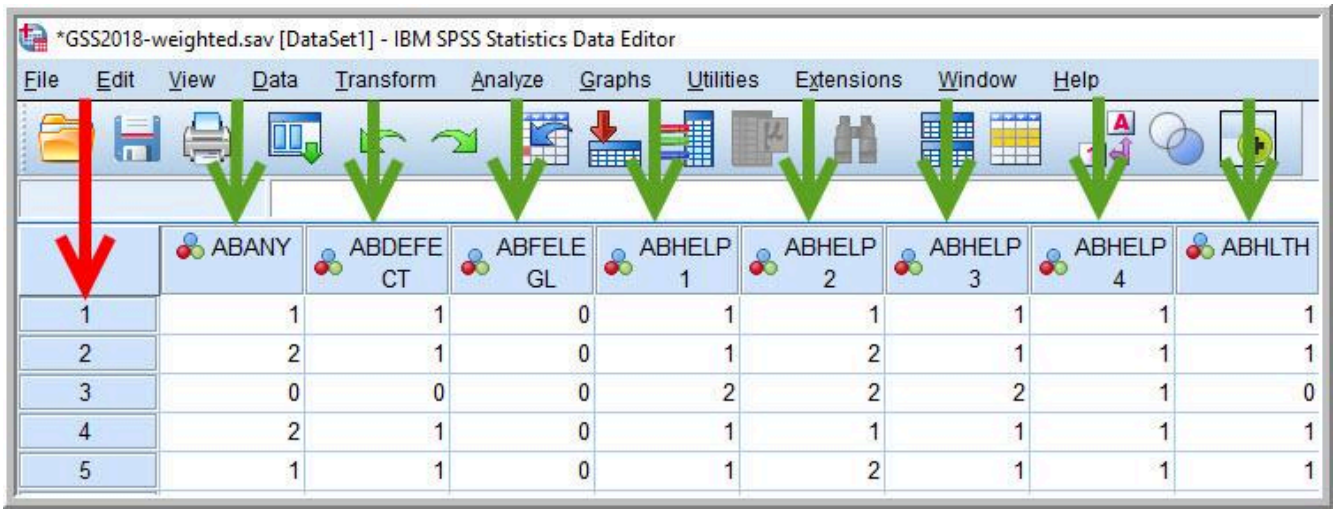
ICPSR_[some number]. You will have to click on multiple folders (e.g., **ICPSR_21600**, then **ICPSR_21600** again, then **DS0002**) to locate the SPSS data file and the PDF of the codebook.

If no option is given for SPSS or whatever software you are using to analyze your dataset, you can click on the **ASCII** link (or the link for another file type), download the file in that format, and then have the program you are using convert the file to its own format.

Viewing a Quantitative Dataset

Regardless of whether you have fielded your own survey or are using data from someone else's survey, at some point you will wind up with a dataset that you will need to analyze. Let's use SPSS to open up a dataset from a national survey with thousands of respondents, the GSS. The GSS is a national probability-sample survey that has been conducted since 1972—at first annually, but only every other year in more recent years because of budget cuts. (You can learn more about how the survey is conducted on its FAQ page.)

Datasets from any year of the GSS can be downloaded free of charge. We will be using the GSS from 2018 for all of the data analysis examples in this chapter. Note that many of the variables we discuss are available in other years as well. If more recent datasets are available, or if you'd like to explore how the attitudes of U.S. adults have changed over time, you can repeat our analyses using the GSS data for other years.



	ABANY	ABDEFECT	ABFELEGL	ABHELP1	ABHELP2	ABHELP3	ABHELP4	ABHLTH
1	1	1	0	1	1	1	1	1
2	2	1	0	1	2	1	1	1
3	0	0	0	2	2	2	1	0
4	2	1	0	1	1	1	1	1
5	1	1	0	1	2	1	1	1

Figure 14.5. Viewing a Dataset in SPSS. A screenshot of one portion of the GSS data file, as seen via the Data View window of SPSS. Five respondents are displayed going down the left side in the rows (as indicated by the red arrow), and eight variables are displayed going across the columns (as indicated by the green arrows).

In SPSS, the GSS dataset you are analyzing will look something like the screenshot in **Figure 14.5**. Note that regardless of the analysis program being used, a typical dataset will be displayed with the units of analysis spread across the rows, and the variables spread across the columns. (Recall that a **unit of analysis** is the type of cases that a researcher is studying, which could be anything from individuals and households, to organizations and countries, to books and movies.) In the GSS dataset, each row represents an individual respondent, and each box (or **cell**) in the dataset represents how that

respondent answered a particular question—specifically, the question measuring the variable listed at the top of the column for that cell.

In the dataset depicted in **Figure 14.5**, we see that Respondent 1 said "1" to the variable *abany* and "1" to *abdefect*. Respondent 2 said "2" to *abany* and "1" to *abdefect*. What do these numbers mean?

Remember that when entering survey responses into a dataset, you always enter them in numerical form—as codes. Furthermore, the questions are listed in the columns under their variable names. To know what questions go with particular variable names, and what response options go with particular codes, you'll need to find documentation describing exactly which questions were asked to collect data for the GSS. You can either examine the codebook provided with the data or go to the Data Explorer on the GSS website, which allows you to look up variables and learn what questions they correspond to and in what years those questions were asked. (Other large surveys have similar online interfaces that can help you identify variables, though most of them force you to pore over their lengthy codebooks.) **Figure 14.6** provides details about the *abany* variable taken from the GSS Data Explorer, and **Figure 14.7** shows the entry in the GSS codebook for the *abdefect* variable.

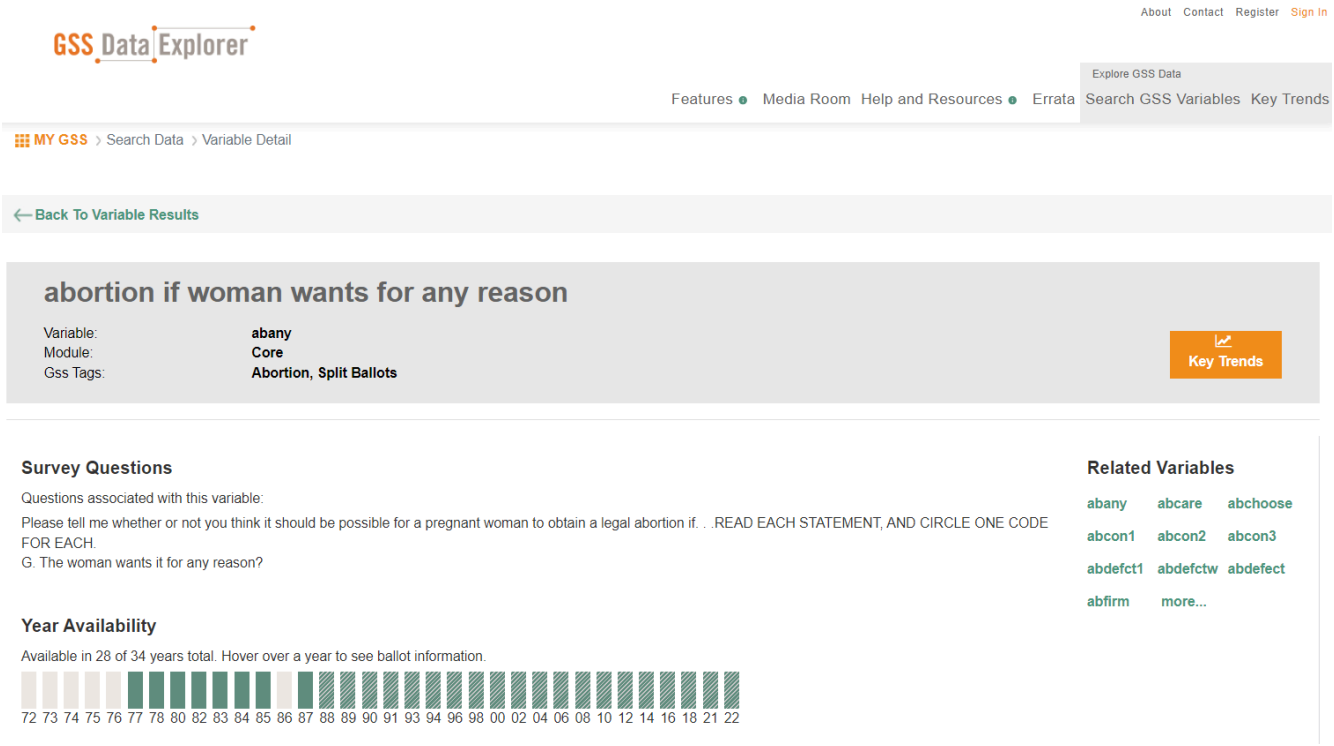
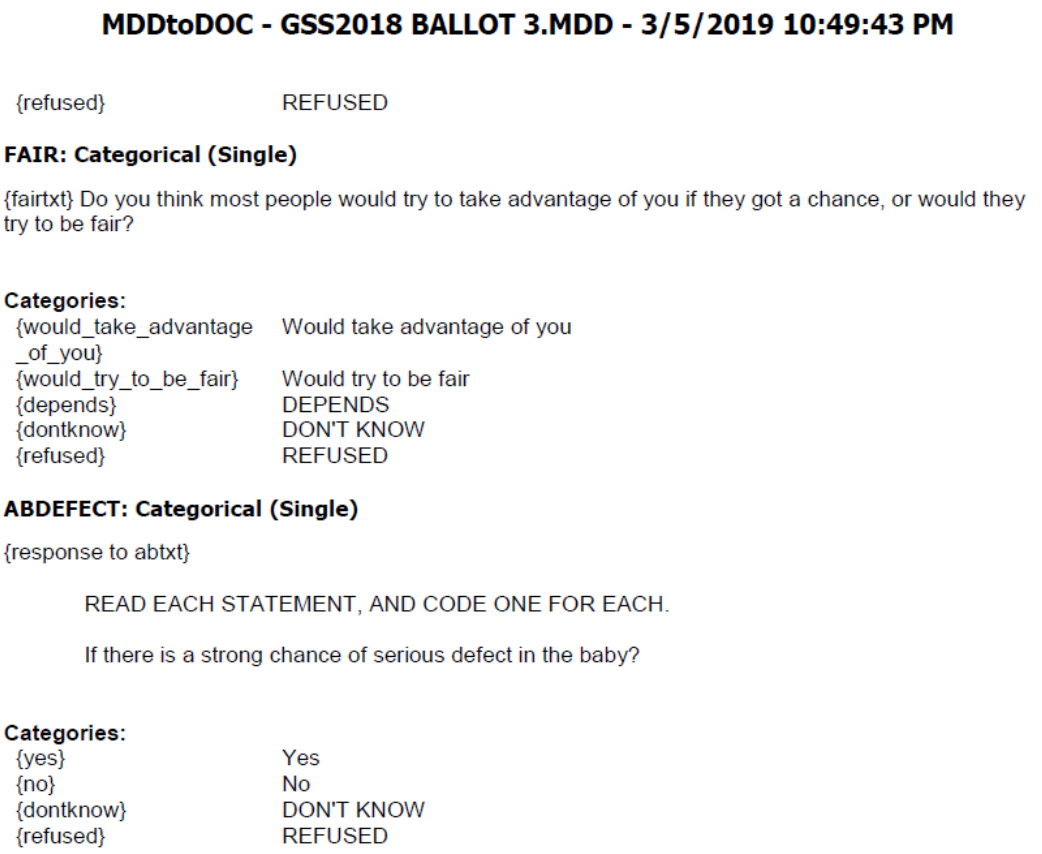


Figure 14.6. Variables in the GSS Data Explorer. The GSS website allows you to search for variables across its decades of datasets and to conduct some basic statistical analyses. The page depicted here lists information about the *abany* variable, including the precise wording of this question about abortion and the survey years when the question was asked. NORC



MDDtoDOC - GSS2018 BALLOT 3.MDD - 3/5/2019 10:49:43 PM

{refused}	REFUSED

FAIR: Categorical (Single)

{fairtxt} Do you think most people would try to take advantage of you if they got a chance, or would they try to be fair?

Categories:

{would_take_advantage_of_you}	Would take advantage of you
{would_try_to_be_fair}	Would try to be fair
{depends}	DEPENDS
{dontknow}	DON'T KNOW
{refused}	REFUSED

ABDEFECT: Categorical (Single)

{response to abtxt}

READ EACH STATEMENT, AND CODE ONE FOR EACH.

If there is a strong chance of serious defect in the baby?

Categories:

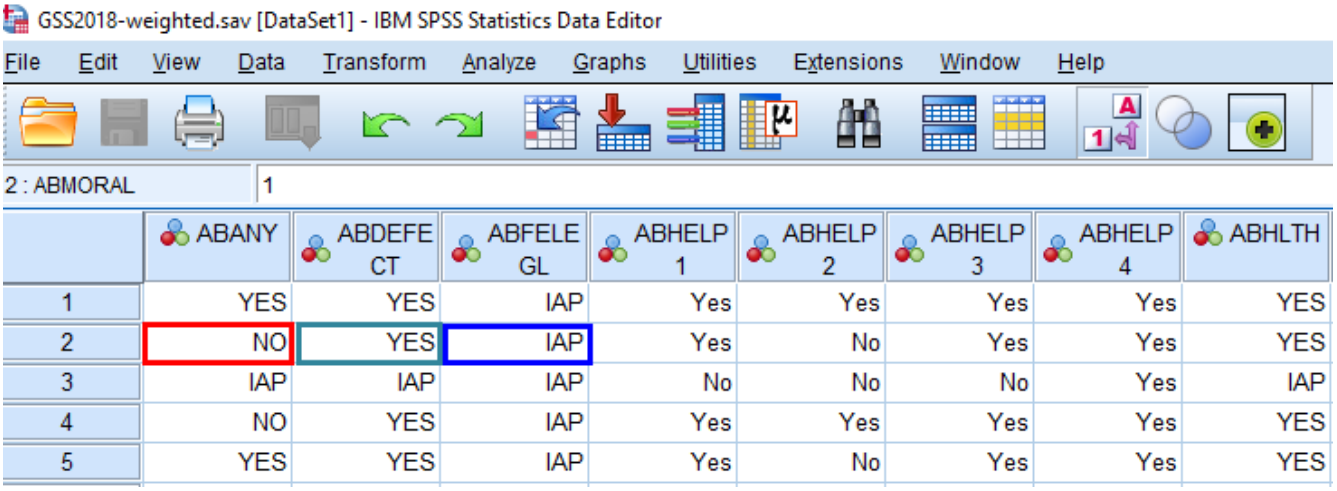
{yes}	Yes
{no}	No
{dontknow}	DON'T KNOW
{refused}	REFUSED

Figure 14.7. Variables in the GSS Codebook. The GSS codebook provides detailed information about the survey's variables, but note that you will need to wade through the computer code (in curly brackets) that the survey's computer system uses to present questions and response options to respondents in the field. The *abdefect* variable is listed under its variable name (ABDEFECT). You can see the part of the question text and all the response options on this particular codebook page, but the introductory part of the question text (used for all the abortion-related questions in the GSS) appears elsewhere in the codebook (you can see that text in Figure 14.5, which features another abortion-related question). *NORC*

Many secondary datasets include *variable labels* that summarize the questions, and *value labels* that tell you what response options the numerical codes reference. In a well-documented dataset, all the variables will be labeled (with variable labels) to represent the questions asked, and all the values of the survey's ordinal- and nominal-level variables will be labeled (with value labels) to represent their response categories. Scale-level variables do not need value labels for their response categories, since the variable values are meaningful by themselves—say, a respondent's age in years, or income in exact dollars.

The GSS dataset includes value labels for its response options, and it also includes variable labels with brief descriptions of each variable, but you should note that these variable labels do not include the entire text of each question. Remember that what exactly a variable measures depends on how exactly its corresponding question was worded, so you should get in the habit of always referring to the full question text rather than going by summaries. Furthermore, many datasets contain very little documentation, so the variable label and value label fields may be blank in your data file. You may want to add or edit these labels in your analysis program, using the questionnaire or codebook as a guide.

In SPSS, you can display the variable labels and other information about a dataset's variables by clicking **Variable View** on the lower left side of the screen. Figure 14.8 shows the variable names in the far left column and information about each variable displayed in each row. Notice that under the **Label** column there's a longer description of the *abany* variable: "Abortion if woman wants for any reason." This variable label helps us a little bit, but we might still be wondering if *abany* refers to whether the respondent has ever had an abortion "for any reason," or whether the respondent supports a woman's right to have an abortion "for any reason." If we look at the survey's documentation (refer to Figure 14.6), we learn that this is the complete question: "Please tell me whether or not you think it should be possible for a pregnant woman to obtain a legal abortion if the woman wants it for any reason?" (Remember: always look up the full text of the question!)



	ABANY	ABDEFECT	ABFELEGL	ABHELP1	ABHELP2	ABHELP3	ABHELP4	ABHLTH
1	YES	YES	IAP	Yes	Yes	Yes	Yes	YES
2	NO	YES	IAP	Yes	No	Yes	Yes	YES
3	IAP	IAP	IAP	No	No	No	Yes	IAP
4	NO	YES	IAP	Yes	Yes	Yes	Yes	YES
5	YES	YES	IAP	Yes	No	Yes	Yes	YES

Figure 14.8. Showing Value Labels for Variables in SPSS. In a survey like the GSS with thorough documentation, the survey creators will provide value labels within the dataset that tell you which response option is associated with which numerical code. Compare this screenshot with the one in Figure 14.5; the numerical codes have been replaced with their corresponding value labels.

Because the nice folks who conduct the GSS also include value labels for response options when appropriate, you can view the datasets with those labels visible—making it easier to determine exactly how a given respondent (the rows in the dataset) responded to a particular question (the columns). When value labels are set, SPSS can instantly swap the numerical codes for each respondent's answers with their corresponding value labels, as shown in Figure 14.8. With these labels displayed, it's easy to see that Respondent 2 (second row) said "NO" to the variable *abany* and "YES" to *abdefect*. Referring to Figure 14.5, which didn't have value labels displayed, we see that the 2 for *albany* means "No" and the 1 for *abdefect* means "Yes."

By the way, the value label "IAP" ("inapplicable") that you see under the *abfelegl* variable means that Respondent 2 was not asked that question. (Note that *abfelegl* is yet another abortion-related question—as tipped off by the letters "AB" starting each of these variable names—and it asks if the respondent thinks "a woman should continue to be able to have an abortion legally or not.") GSS employs a **split-ballot design**, which splits the overall sample into groups and uses a different set of supplementary questions for each group. In this case, the questionnaire (or ballot) that Respondent 2 received did not include the *abfelegl* question, so this person's answer was marked as IAP. As you can see by referring to Figure 14.5, IAP was numerically coded as 0.

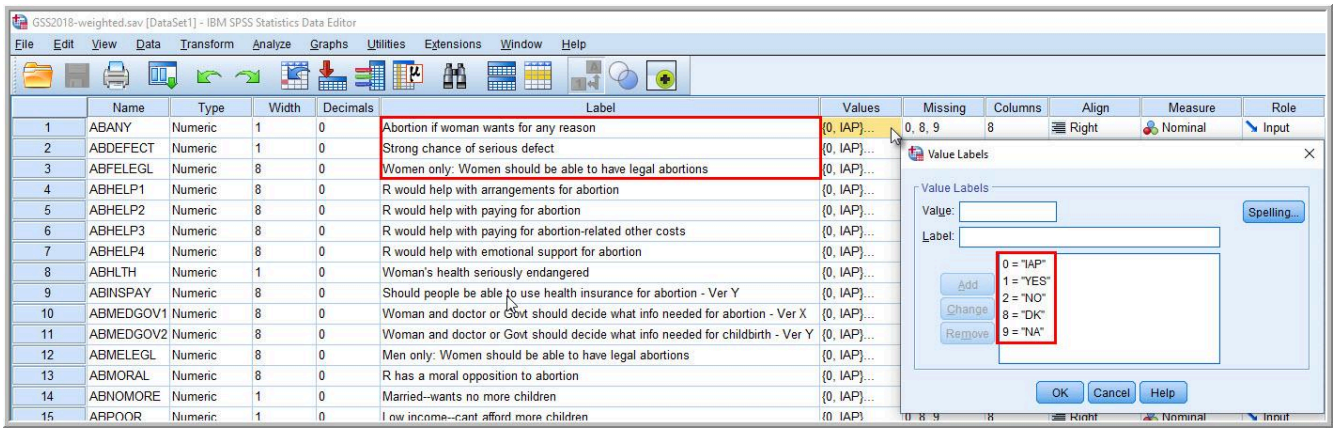


Figure 14.9. Listing the Value Labels for a Variable. From the SPSS Variable View window, you can open up a list of all the values for a given variable—assuming that the creators of the dataset labeled them. Click on the cell for that variable underneath the **Values** column. On each row of the list, you will see the numerical code followed by its corresponding value label. As described in the text, we need to look in another column of the Variable View window (the **Missing** column) to learn which of these values are missing values.

In a well-documented dataset, all the numerical codes for a particular variable will be listed, along with their corresponding value labels, in that variable's entry in the SPSS **Variable View** window. (Make sure to click the tab on the bottom if you are currently in the **Data View** window.) If we click on the **Values** cell for a variable and then click on the three dots that appear on the right-hand side of the cell, a window opens up with the details for each response option. As **Figure 14.9** shows, now we can see the numerical codes and associated value labels for *abany* in one place:

- 0 (IAP for "Inapplicable")
- 1 (YES)
- 2 (NO)
- 8 (DK for "Don't know")
- 9 (NA for "No answer")

To the right of the **Values** column in the Variable View window, there is a **Missing** column. By clicking on the cell under that column, we learn that *albany* has been assigned three missing values: 0, 8, and 9. These correspond to IAP ("Inapplicable"), DK ("Don't know"), and NA ("No answer"), respectively. Since these three values are missing values, the other two values—1 (YES) and 2 (NO)—are our valid values for this variable. As we'll describe in the next section, only valid values will appear in our analysis, and missing values will be excluded—which is what we want.

Key Takeaways

1. Among other things, a codebook tells the researcher the variable name that is assigned to each survey question and the numerical values assigned to each response category. Before you start entering data you

collected in SPSS, you need to create your own codebook that includes these details. Then you can enter the responses that each of your respondents provided as numerical codes, which is the best approach to data entry.

2. You can easily find secondary data by searching the web or online data repositories that store many datasets. Using secondary data will allow you to skip the data collection process, and if you're lucky, you can locate a dataset that asked the specific questions you are interested in and that obtained a representative sample of your population of interest.
3. Using a statistical analysis program like SPSS, you can view your data as numerical codes or with value labels, and you can see details about each of your variables, including whatever labeling has been provided by the survey's creators.

14.2. Univariate Data Analysis: Analyzing One Variable at a Time

Learning Objective

Understand the various ways that you can conduct a univariate analysis using tables and charts, and how those approaches differ depending on your variable's level of measurement.

When you start analyzing a dataset, it's a good idea to spend some time getting to know the variables that will be central to your analysis. This involves analyzing them individually—what we call **univariate analysis** (i.e., an analysis of one variable at a time). While univariate analysis is rarely the goal of academic research, we engage in it all the time to describe the characteristics of our samples—in the process, generating what we call **descriptive statistics**. We want to know, for instance, what the people we interviewed are like in terms of their personal background and social identities. We want to know what they think about the topics we addressed in our survey. Simple univariate analyses can tell us a great deal about these things, and if we are lucky enough to have a representative sample, we should be able to generalize our results to our target population.

For **categorical variables** (i.e., variables measured at the nominal or ordinal level), conducting a univariate analysis typically involves calculating a **frequency distribution**. Figure 14.10 features the resulting table (a *frequency table*) for a GSS variable called *pray* that measures how frequently a respondent prays.

A	B	C Frequency	D Percent	E Valid Percent	F Cumulative Percent
Valid	Never	370	15.8	15.9	15.9
	Less than once a week	251	10.7	10.7	26.6
	Once a week	152	6.5	6.5	33.1
	Several times a week	231	9.8	9.9	43.0
	Once a day	650	27.7	27.8	70.9
	Several times a day	680	29.0	29.1	100.0
	Total	2334	99.4	100.0	
Missing	DK	2	.1		
	NA	11	.5		
	Total	14	.6		
Total		2348	100.0		

Figure 14.10. Breaking Down a Frequency Table. In this frequency table generated in SPSS, each column provides different information about the distribution of values for the variable *pray* (which measures how often the respondent prays). Refer to the discussion in the text about what each column refers to.

Starting at the far left side of the table shown in**Figure 14.10**, here's what the columns in the frequency table mean:

A: Valid/Missing/Total: Column A describes the valid values, the missing values (here, DK for "Don't know" and NA for "No answer"), and the totals for all cases, respectively.

B: Labels: Column B provides brief descriptions of the attributes of the variable (i.e., the response options).

C: Frequency: Column C lists the number (or **count**) of respondents giving each possible response, including any nonresponses designated as missing values.

D: Percent: Column D lists the percentage of respondents giving each possible response. This percentage includes missing values—which you typically are not interested in.

E: Valid Percent: Column E lists the percentage of respondents giving each possible re**valid percentage**sponse that has been designated a valid value (i.e., not a missing value). The valid percentage is the number you normally want to use when discussing results.

F: Cumulative Percent: Column F lists the percentage of respondents giving each valid response added to all of the percentages coming before that one. For example, 73.3 percent of respondents said they pray once a week or more often than that. This percentage was calculated by adding the "Valid Percent" figures from "Once a week" to "Several times a day."

When you write up your results in a report, you should summarize for readers the key patterns in your tables. Consider the following two statements that you might make about the frequency table shown in **Figure 14.10** to convey its findings:

1. **"Most respondents report praying frequently. Specifically, 57 percent pray at least once a day."** We combined two categories here: 28 percent of respondents pray "once a day," and 29 percent pray "several times a day," which together equal 57 percent. Note that we also rounded the percentages in the table to the nearest whole percent. This clarifies the numbers for readers, and we suggest doing this unless you have many figures that, when rounded, would be the same (which then wouldn't allow you to see any differences between them).
2. **"The next most prevalent categories are those who pray less than once a week (11 percent) and those who never pray (16 percent)."** In this statement, we are noting additional categories in order of their size. Unless there's a compelling reason to highlight the smallest groups, it's okay not to mention them—the reader will just assume the remaining respondents fall in those categories. Alternatively, we could consolidate all the (small) middle categories, which total to 27 percent. For theoretical reasons, you'd probably want to keep the respondents who say they "never" pray as its own category rather than consolidating it: this group is likely to be quite different from, say, those who pray but do so infrequently (for one thing, atheists would fall into the former group but not the latter group).

In your written interpretation, you could mention the results for every single response. We recommend you don't do that *once your response categories number four or more*. Instead, combine smaller categories (as we did in the first statement), or leave them out (as we suggested was possible for the second statement).

Frequency distributions for nominal- and ordinal-level variables can be represented visually using a **pie chart** or **bar chart**. As shown in **Figure 14.11**, a pie chart illustrates the valid responses by taking 100 percent as representing the whole pie and giving each value a slice of the pie that matches its portion of the total. This pie chart provides a visualization of the data in our earlier frequency table that reported how often respondents prayed.

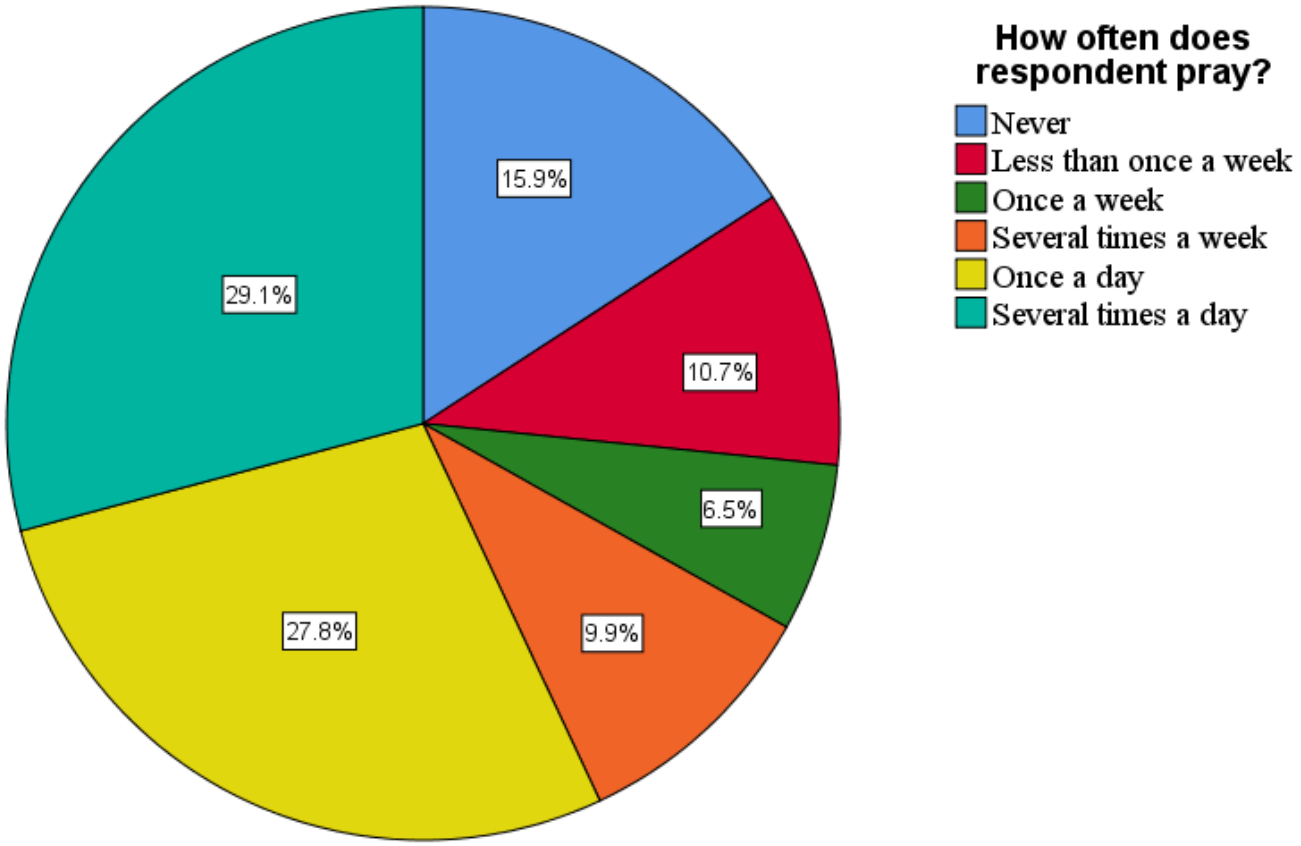


Figure 14.11. Pie Chart for Frequency of Prayer. A pie chart represents all values of a variable within one circle, with the slices proportional in size to the corresponding frequencies of each valid response. In this SPSS-generated pie chart, the colors of the chart labels correspond to the value labels of the variable (listed in the chart's legend). Some social scientists look down on the humble pie chart, whose design can make it harder to compare similarly sized categories; they tend to favor its sexier cousin, the bar chart.

Without valid percentages displayed in a pie chart (which we added in **Figure 14.11**), it is sometimes difficult to say whether one slice is bigger than another similarly sized one. In this figure, for instance, the slices for "Several times a day" and "Once a day" are fairly close in size. This happens frequently with pie charts, especially if the categorical variable we're using has more than three values. For this reason, social scientists tend to prefer bar charts over pie charts. A *bar chart* illustrates the frequency distribution by showing the valid values of a variable as vertical or horizontal bars. **Figure 14.12** depicts a bar chart for the prayer variable we analyzed earlier.

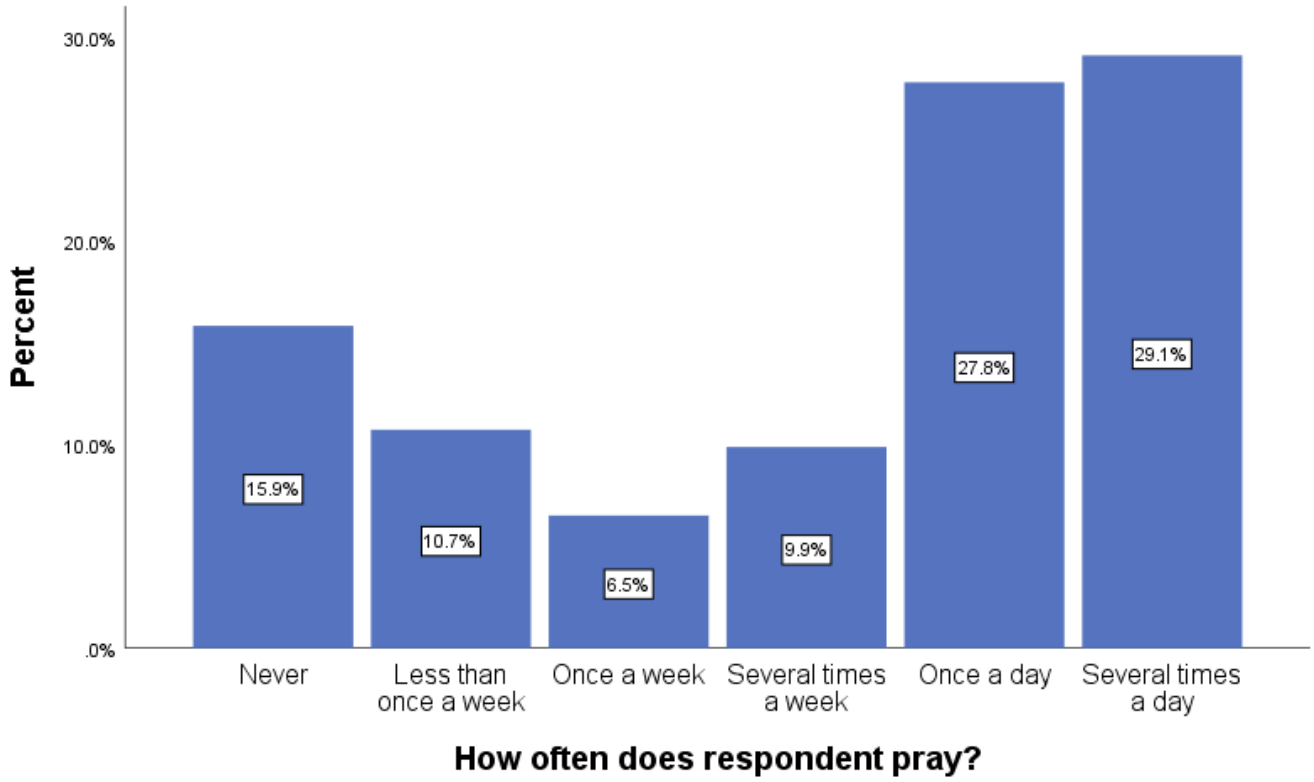


Figure 14.12. Bar Chart for Frequency of Prayer. A bar chart compares different response options for a categorical variable based on their counts or percentages. The height of each bar (or the length in a horizontal bar chart, which is also perfectly acceptable) reflects the proportion of respondents who answered the question in that way. By default, many statistical analysis programs (including SPSS) generate charts with counts (the number of respondents), rather than percentages (the proportion of all respondents). You should change the settings to display percentages. Researchers tend to prefer percentages on their tables and charts, given that they are usually interested in the proportions of each category in the target population, not the counts in their specific sample.

Even if we hadn't shown the valid percentages in the bar chart, it would be easy to tell whether "Several times a day" or "Once a day" was the larger category, since "Several times a day" is taller. In a bar chart, each bar indicates a value's relative frequency, and the side-by-side comparison makes determining which value is larger a piece of cake (to continue with our baked-goods motif).

We can also use scale-level variables in frequency distributions and charts. For example, age is a scale-level variable.[1] Its values in our GSS dataset range from 18 to 89. (Note that the distribution starts at age 18 because the GSS collects data only from adults.) That's a lot of values to show in a frequency table! Luckily, we can use data analysis programs to take the numerous values of a scale-level variable and place them into a smaller number of buckets—which become the categories of our new variable. This allows us to generate a frequency table that's much easier to understand.

1. The age variable in the GSS is scale level from 18 to 88 (with one year separating each value), but the highest value possible is 89–which represents not only those age 89 but also any older respondents. For that reason, the GSS age variable is actually an ordinal-level variable. That said, we can treat it as scale level for practical purposes, as we do in this chapter.

In the table depicted in **Figure 14.13**, we've *collapsed* (i.e., combined) the 72 different values of our age variable into just three categories. (The minimum number of values for a variable is two, since there has to be some "variation" in the attributes of a "variable.") Essentially, we've changed the codes in our variable, moving from a range of 18 to 89 to—for instance—1 to 3 (if we want to use consecutive numbering for our three ordinal categories). We call this process **recoding**. Social scientists recode variables all the time to highlight certain features of the data or simplify their analysis. But note that by recoding the variable and collapsing its original categories, we have sacrificed detail in our data. Specifically, the new variable will tell us only where a person falls within a range—not what their specific age is. For this reason, you typically want to create a new variable rather than replacing the original variable with its recoded structure. That way, you preserve the detail in the original variable.

Age

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	18-39	934	39.8	39.9	39.9
	40-64	978	41.6	41.8	81.7
	65 and older	428	18.2	18.3	100.0
	Total	2340	99.6	100.0	
Missing	NA	8	.4		
Total		2348	100.0		

Figure 14.13. Frequency Distribution for a Recoded Age Variable. If we tried to generate a frequency table for a scale-level variable like age, we would have a table with dozens of rows—one for each possible value of age in the sample. A better approach is to *collapse* the age variable into a new variable with just three categories: here, ages 18 to 39, ages 40 to 64, and age 65 and older.

How would we summarize the results of this frequency table? As we did with the frequency table shown in **Figure 14.10**, we probably want to start by saying which response category has the highest percentage of respondents, and then move to the next-largest category, and then the smallest. We might say something like this: "Within the GSS sample of U.S. adults, 42 percent of respondents are between the ages of 40 and 64, and 40 percent are between the ages of 18 and 39. The oldest age-group, those 65 and older, is smaller than the other two (18 percent)."

What if we want to present our frequency distribution for the age variable in a chart? Pie and bar charts are not good options for scale-level variables. For instance, our age variable would have 72 different bars if we created a bar chart for it. We would not be able to label each value properly. Of course, we could create a bar chart with our recoded, ordinal-level age variable, but we can use two other types of charts specifically for scale-level variables: *histograms* and *line graphs*.

In a **histogram**, the height of each bar represents the frequency of the variable values shown, just like in a bar chart. Unlike a bar chart, however, a histogram does not show gaps between the bars, which makes it easy to visualize the shape of the distribution (see the example in **Figure 14.14**). You wouldn't bother labeling values in a histogram; instead, you want to see the general trend across values of the variable. Note that data analysis programs like SPSS allow you to change the width of the bars in a histogram (i.e., the size of the *bins*). You might want to experiment with the size of the bins in your histogram to see if different sizes give you a different sense of the shape of the distribution.

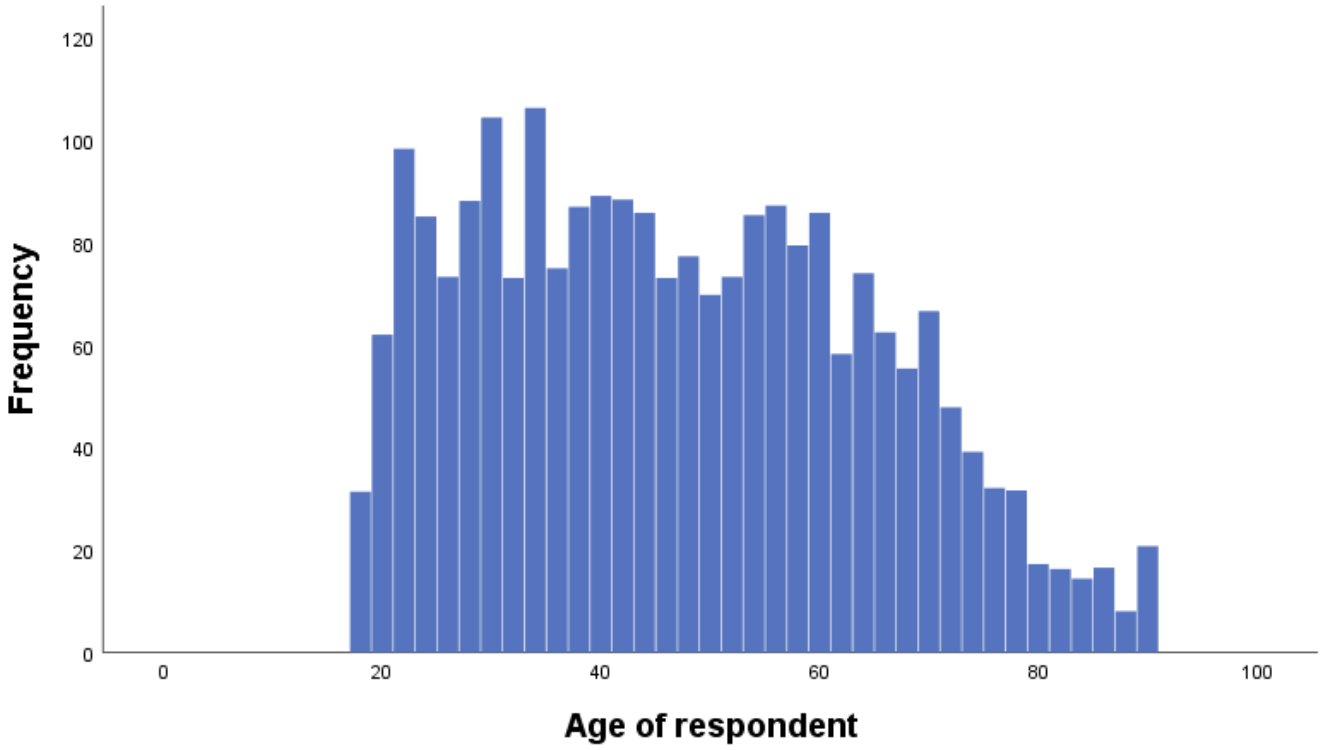


Figure 14.14. Histogram of the Scale-Level Age Variable. Bar charts and pie charts are not appropriate for scale-level variables like this age variable from the GSS. Instead, we should use a *histogram*, which removes the spaces between the bars so that we can clearly see the shape of the distribution.

A **line graph** is another option for scale-level variables. Instead of bars, the height of a continuous line represents the frequency of the values of the variable. As you can see in **Figure 14.15**, a line graph follows the peaks and valleys of the histogram. (The upturn at the end of the histogram and line graph is due to the fact that the highest possible value for age in GSS 2018 is "89 or older." Thus, the value contains both respondents who are 89 years old *and* older respondents.)

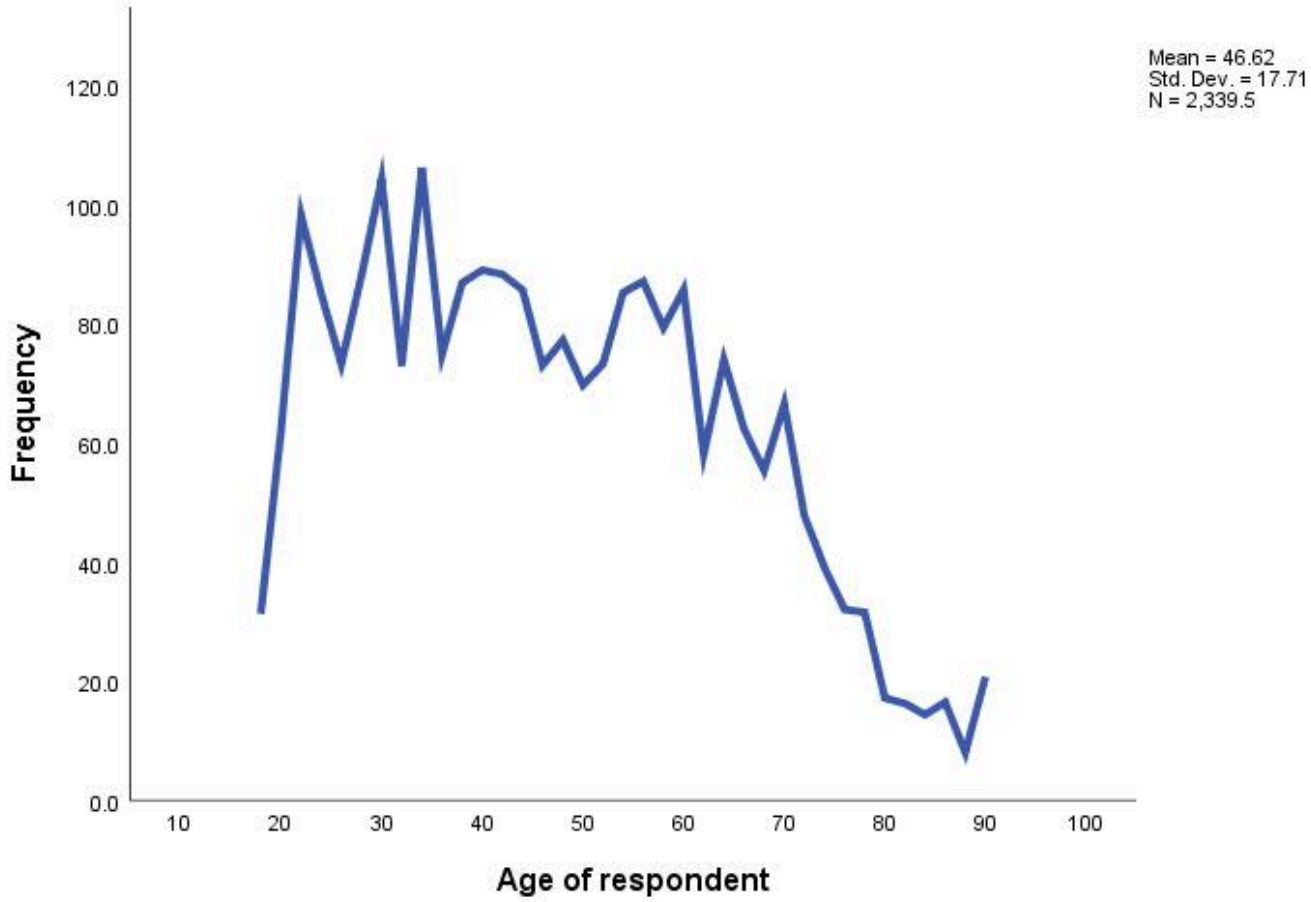


Figure 14.15. Line Graph of the Scale-Level Age Variable. Another chart option for scale-level variables is a *line graph*. It looks similar to the histogram, with the line drawn by the data analysis program sketching an outline of the histogram bars—allowing you to see the shape of the distribution rather clearly.

Measures of Central Tendency



Sociologists frequently report two *measures of central tendency*—the *median* and *mean*—for variables in their analysis. The shape of the underlying frequency distribution determines where the median falls relative to the mean. For instance, a *left-skewed distribution*—where many people have high scores on the variable, and a few people have low scores—will have a mean that is lower than the median. Meanwhile, a *right-skewed distribution*—where many people have low scores on the variable, and a few people have high scores—will have a mean that is higher than the median. The income distribution in most countries is right-skewed: most people make low or middling incomes, but a small group of billionaires make much, much more. For this reason, government agencies routinely report the median household income and not the mean. Given the shape of the country's income distribution (thank you, Elon Musk), the median is a better measure of what an "ordinary" household makes. *Andrea Piacquadio, via Pexels*

When conducting univariate analyses to describe their samples, researchers frequently calculate statistics that summarize key characteristics of a variable's frequency distribution. In this section and the next, we'll look at the two most common—*measures of central tendency* and *measures of variability*. **Measures of central tendency** give us information about the typical value in a frequency distribution. The most commonly used measures of central tendency are the **mode**, the **median**, and the **mean**:

Mode: The value that occurs most frequently in the distribution of a variable.

Median: The value that comes closest to splitting the distribution of a variable in half. In other words, 50 percent of cases are below the median and 50 percent are above the median.

Mean: The arithmetic average of the distribution of a variable. The mean is obtained by adding the values of all the cases and then dividing by the total number of cases.

It's up to you as a researcher to decide which measures of central tendency to include in your report. Important factors in this decision include (1) the level of measurement of the variable you are analyzing; (2) the nature of your research question; and (3) the shape of the distribution. We will discuss these considerations in turn.

The simplest measure of central tendency is the *mode*, the most frequently occurring value. You can calculate the mode for variables across all levels of measurement, and it is easy to spot when you are looking at a frequency distribution in a chart or table—it's the highest bar in a bar chart, and it's the cell with the highest frequency (count or percentage) in a frequency table. In the frequency table depicted in **Figure 14.16**, the modes for several GSS variables are shown in red boxes. Note that we have three different variables in this frequency table—each of which has its own mode—but we are still analyzing the variables one at a time (univariate analysis) rather than assessing the relationship between them (bivariate or multivariate analysis).

		%	N
Gender	Male	45.5%	1069
	Female	54.5%	1279
	Total	100.0%	2348
Race identification	Whites	66.7%	1553
	Nonwhites	33.3%	776
	Total	100.0%	2328
How often respondent attends religious services	Never	29.8%	694
	Less than once a year	5.6%	130
	Once a year	12.9%	302
	Several times a year	10.0%	233
	Once a month	6.4%	149
	2-3 times a month	8.5%	198
	Nearly every week	3.9%	91
	Every week	17.3%	403
	More than once a week	5.6%	131
	Total	100.0%	2331

Figure 14.16. Three Frequency Distributions with Their Modes Highlighted. This frequency table shows the frequency distributions for three variables in the GSS: the gender of the survey's respondents, whether or not the respondents were white, and how often the respondents attend religious services. The mode for each variable (the red boxes) can be easily determined by identifying the highest count or percentage among each variable's categories.

Determining the mode becomes more complicated, of course, if we are analyzing scale-level variables with a large number of values. For instance, we could generate a (very lengthy) frequency table for our age variable, but then we would have to go through its many rows to find the modal value. Fortunately, data analysis programs can calculate the mode for us, as illustrated in **Figure 14.17**.

When writing up these results, note that you never need to report the numerical code associated with a categorical variable, even if you are reporting the mode. For instance, for a nominal-level variable like gender, do not report that the mode is "2"; instead, identify the mode as the response option that "2" refers to, which for the GSS gender variable is "female." With scale-level variables, however, we can refer to the numerical values directly: the mode for the age variable is 34 years, and the mode for the educational variable is 12 years, as depicted in **Figure 14.17**. Note that if two or more values for a given variable have the same frequencies, the distribution will have multiple modes, but SPSS will show only one mode in the table it generates.

Statistics

		Age of respondent	Highest year of school completed
N	Valid	2340	2345
	Missing	8	3
Mode		34	12

Figure 14.17. Modes for Age and Educational Attainment. Rather than generating a frequency table that lists all the values of scale-level variables like age and educational attainment, we can ask SPSS (or whatever data analysis program we're using) to calculate the modal values for us.

The *median*, or middle value in a distribution, can be used with *ordinal- and scale-level variables only*. That's because you've got to be able to rank the cases from low to high (or high to low) to pick a middle case. SPSS can calculate the median in its frequency tables, as shown in **Figure 14.18** for

the GSS frequency of prayer variable. The program identifies the median as the value that comes as close as possible to splitting the distribution into two equal halves.

The median can also be thought of as the 50th **percentile**—the point at which 50 percent of the values in a distribution are at that level or below it. Therefore, in the "Cumulative Percent" column of an SPSS frequency table, the median value would be the value at which the 50th percentile is equaled or exceeded (one reason that this column is actually useful). For instance, you can see in Figure 14.18 that the **cumulative percentage** rises above 50 on the row for "Once a day," reaching 70.9. This tells us that "Once a day" is the median value.

Statistics

How often does respondent pray?

N	Valid	2334
	Missing	14
Median		5.0000

How often does respondent pray?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00 Never	370	15.8	15.9	15.9
	2.00 Less than once a week	251	10.7	10.7	26.6
	3.00 Once a week	152	6.5	6.5	33.1
	4.00 Several times a week	231	9.8	9.9	43.0
	5.00 Once a day	650	27.7	27.8	70.9
	6.00 Several times a day	680	29.0	29.1	100.0
	Total	2334	99.4	100.0	
Missing	System	14	.6		
Total		2348	100.0		

Figure 14.18. Median for Frequency of Prayer. By default, the frequency tables that SPSS generates do not tell us what the median value for each variable is. However, we can identify the median by referring to the **Cumulative Percent** column: the median value is the row where the cumulative percentage first exceeds 50, which in this example is the row for the "Once a day" response option. (This technique works because the median is equivalent to the 50th percentile—the point at which 50 percent of values are at that level or below it.) SPSS can also calculate the median for us if we convert our variable into a scale-level variable, as in the example; the statistics table at the top left indicates that the median is "5.00," which is the numerical code assigned to "Once a day."

It's clear that the median for our example—"Once a day"—does not actually split the distribution into two equal parts, but it's as close as we can get using the categories of this ordinal-level variable. For ordinal variables with relatively few values, the median will probably not be a precise measure of central tendency.

The *mean*, or arithmetic average, is most appropriately used with scale-level variables. For example, the mean age in our

GSS sample is 46.6 years, and the mean number of years of education is 13.7 years. Whether we can calculate a mean for ordinal-level variables is a matter of some social scientific controversy. Statistical purists point out that the values of an ordinal-level variable do not specify the distances between those values, which makes it impossible to determine a mean. Nevertheless, many social scientists dutifully report means for ordinal-level variables, arguing that this measure provides more information than the median or mode do about the distribution of a particular variable. We fall in the latter camp, but we advise you to be cautious when interpreting the mean for an ordinal-level variable. See the sidebar **Deeper Dive: Treating Ordinal-Level Variables as Scale-Level Variables** for more on this topic.

Table 14.1 summarizes which measures of central tendency we can calculate based on a variable's level of measurement.

Table 14.1. Measures of Central Tendency for Variables at the Scale, Ordinal, and Nominal Level

Measurement Level	Example Variables	Mode	Median	Mean
Scale (interval or ratio level)	Age, educational level	Yes	Yes	Yes
Ordinal	Frequency of attendance at religious services, frequency of prayer	Yes	Yes	With caution!
Nominal	Gender, race identification	Yes	No	No

Note that for a scale-level variable, you should also consider the shape of the distribution. In a perfectly symmetric distribution, the mean, median, and mode will all be equal. As shown in the right-hand histogram in **Figure 14.19**, a distribution with a tail to the right (*right-skewed*) will have a mean that is higher than the median. As shown in the left-hand histogram, a distribution with a tail to the left (*left-skewed*) will have a mean that is lower than the median.

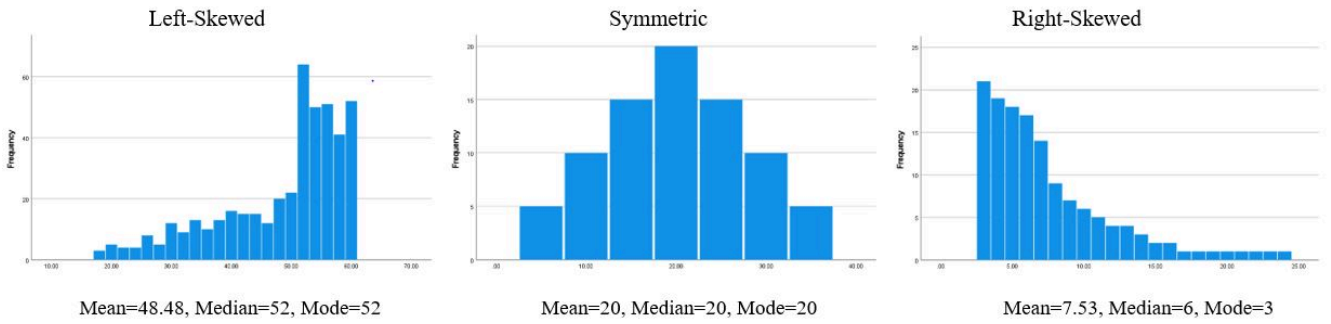


Figure 14.19. Histograms Showing Shapes of Frequency Distributions. To get a better sense of our data, it is a good idea to graph scale-level data in a histogram and to examine the shape of the frequency distribution. The mode, median, and mean will be fairly similar in a perfectly symmetrical distribution. However, for a left-skewed distribution, the median will have a higher value than the mean. For a right-skewed distribution, the opposite is true.

The median will be a better measure of central tendency than the mean to use with a very skewed distribution. For example, if we are analyzing income levels in a national sample of U.S. households, the small number of super-rich

people will pull the mean much higher than the median (a *right-skewed distribution*). That's why published U.S. Census Bureau figures report the median income instead of the mean income. Median income is a better measure of what the earnings of an "ordinary" American household look like.

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://viva.pressbooks.pub/sociology-research-methods/?p=495#oembed-1

Video 14.5. Measures of Central Tendency. Here's a good review video for measures of central tendency.

Deeper Dive: Treating Ordinal-Level Variables as Scale-Level Variables

Let's consider an example to illustrate the benefits and pitfalls of calculating means for ordinal-level variables. **Figure 14.20** shows an SPSS statistics table with the mean score for the frequency of prayer variable.[2]

Even though SPSS spits out this calculation on command, you can't just report the mean frequency of prayer as "4.11." That's because the numbers for ordinal-level variables are somewhat arbitrary. Yes, they have to be in order from low to high or high to low, but we don't know what the numbers mean unless we know which value labels go with which numbers. Let's review this variable's numerical codes and its associated response categories:

1 = Never

2 = Less than once a week

3 = Once a week

4 = Several times a week

5 = Once a day

6 = Several times a day

Calculating the mean for this variable assumes that these numerical codes from 1 to 6 are meaningful as numbers. But as you can see, the distances between each of these response options are not precisely

Statistics

How often does respondent pray?

N	Valid	2334
	Missing	14
Mean		4.1055

Figure 14.20. Mean for Frequency of Prayer, an Ordinal-Level Variable. We can convert our ordinal-level variables into scale-level variables and then calculate a numerical mean, which would be the average of the numerical values for each response recorded in our sample. We should be careful, however, about how to interpret a numerical score for an ordinal-level variable.

2. The original numerical coding for the GSS frequency of prayer variable used 1 for "Several times a day" and 6 for "Never." For this example, we've reversed the ordering of the response options so that 6 now indicates "Several times a day" and 1 indicates "Never." We've recoded the variable in this way to make it more intuitive, with higher numerical values signifying a higher frequency of prayer.

determined and not equivalent. “Several” (in “Several times a week” and “Several times a day”) is vague. The distances between 4 (“Several times a week”) and 5 (“Once a day”), and between 5 (“Once a day”) and 6 (“Several times a day”), are both 1 “unit,” but it’s not clear what that “unit” really means, or if moving between these three categories truly covers the same distance. And what does the variable’s mean (4.11) represent within this range of values? Specifically, what does moving 0.11 units up the scale—from 4 toward 5—really mean? We don’t know, and therefore we can’t really say much about a value like 4.11—except that the mean frequency of prayer is somewhere between “Several times a week” (coded as 4) and “Once a day” (coded as 5).

In general, we would say that the mean can be used with ordinal-level variables with the caution that it must be interpreted in light of the value labels below and above that number. We can say the mean falls somewhere between two values, but we can’t say anything more than that. Whether it’s even useful to calculate the mean, then, is a judgment call.

We would say that the means for four- or five-point Likert scales and similar ranges of responses are a bit more intuitive. That’s because the Likert scale responses (on a four-point scale, “Strongly disagree,” “Disagree,” “Agree,” “Strongly agree”) are balanced on either side, and the responses appear to move steadily from lesser to greater intensity (of agreement or disagreement, depending on the scale). Therefore, it is easier to assume that moving from “Strongly disagree” to “Disagree” is equivalent to moving from “Strongly agree” to “Agree,” or that a 0.5 difference between a value of 1 and 1.5 is something similar to a 0.5 difference between a value of 2 and 2.5.

In any case, you will frequently see social scientists calculating means for Likert-scale variables. Sometimes, ordinal-level variables will even be used in multivariate analyses that apply techniques appropriate only for scale-level variables. You should use your own judgment about whether those uses are justified, keeping the pitfalls we’ve mentioned in mind.

Measures of Variability



If your friend jumped off a cliff, would you? Hopefully, only if you have measured both the central tendency and variability of the water's depth in that area. *Kammeran Gonzalez-Keola, via Pexels*

Let's say you and some friends are on top of a cliff overlooking the ocean and are deciding whether or not to jump into the water. One of your friends has researched the beach and learned that the average depth of the ocean at this point is 20 feet—clearly, deep enough for a cannonball dive. Does that give you sufficient information to hurl yourself into the ocean?

No. Even if 20 feet is the mean, the ocean depth could range from, say, 3 to 30 feet. In other words, there's likely to be *variability* around the mean score. The ocean at this point on the shoreline probably isn't 20 feet deep everywhere. For the sake of that beautiful social scientist brain of yours, at least do more research before you leap.

Getting a good sense of your data requires you to go well beyond the usual statistical suspects—mean and median—and understand how much your data points vary or stay consistent. Common **measures of variability** include the **range**, **interquartile range**, **standard deviation**, and **index of qualitative variation**:

Range: The distance between the lowest and highest values in a distribution.

Interquartile range (IQR): The distance between the values at the 25th percentile and the 75th percentile in a

distribution. The 25th percentile is the point at which 25 percent of the values in a distribution are at that level or below it. The 75th percentile is the point at which 75 percent of the values in a distribution are at that level or below it. In other words, the IQR spans the middle 50 percent of a distribution.

Standard deviation (SD): A measure of the average distance of all scores from the mean score. This calculation takes every single value in the distribution into consideration.

Index of qualitative variation (IQV): A measure of how much the distribution varies from one in which all cases are concentrated in one category.

As for measures of central tendency, which measures of variability are appropriate to use depends on the measurement level of the variable you are analyzing, as indicated in **Table 14.2**.

Table 14.2. Measures of Variability for Variables at the Scale, Ordinal, and Nominal Level

Measurement Level	Variable Examples	Index of Qualitative Variation	Range	Interquartile Range	Standard Deviation
Scale (interval or ratio level)	Age, educational level	No	Yes	Yes	Yes
Ordinal	Frequency of attendance at religious services, frequency of prayer	No	Yes	Yes	With caution!
Nominal	Gender, race identification	Yes	No	No	No

Table 14.3 breaks down the measures of variability for some of the GSS variables we've been using. The first two variables, *age of respondent* and *highest year of school completed*, are scale-level, so their figures can be interpreted with no additional information. For example, the highest number of years of school completed is 20, and the lowest is zero (0), so the range is highest minus lowest, which equals 20 years.[3] Although the range is easy to calculate, exceptionally high or low numbers (called **outliers**) can inflate it considerably. For example, if only one person in the GSS sample had completed 35 years of formal schooling, the range would be 35 years, which does not convey the narrower variability of most people's education in the sample. (To a lesser extent, outliers can also affect the standard deviation of a distribution.)

3. Like age, the GSS variable for years of formal education combines the highest values into one value (20 or more years). Therefore, it is not a true scale-level variable, but it can be treated as one for most analyses.

Table 14.3. Range, IQR, and Standard Deviation for Selected GSS Variables

Variables	Range	Interquartile Range	Standard Deviation
Age of respondent	71 (18–89)	28 (32–60)	17.71
Highest year of school completed	20 (0–20)	4 (12–16)	3.02
How often the respondent attends religious services	8 (0–8)	6 (0–6)	2.82
How often does the respondent pray?	5 (1–6)	4 (2–6)	1.83
Respondent's highest degree	4 (0–4)	2 (1–3)	1.22

Measures of variability for ordinal-level variables are more complicated to interpret. For instance, our variable for attending religious services goes from 0 to 8, but a range of "8" is meaningless without knowing the actual response options the numerical codes refer to. (As it turns out, 0 means "Never" and 8 means "Several times a week.") The IQR is based on percentiles, so we run into the same issue with ordinal-level variables when we calculate it as we did previously with the median. Although the program will do your bidding and output numbers for these measures of variability, the results won't be very helpful in understanding the nature of an ordinal-level variable.

The index of qualitative variation (IQV) is the only measure of variability appropriate for nominal-level variables.[4] When expressed as a percentage, as in **Table 14.4**, the IQV can vary between 0 and 100 percent. The closer to 0 percent the IQV is, the higher the percentage of cases in just one category—until ultimately all cases are in a single category, indicating no variation at all in the distribution. The closer to 100 percent, the more the cases are distributed evenly across the categories—until ultimately each one has an equal number of cases, indicating maximum variation.

4. SPSS does not directly calculate the IQV. However, you can find calculators on the internet or enter the formula into a spreadsheet.

Table 14.4. Frequency Distribution and Index of Qualitative Variation for Gender and Race

	%	N	IQV	Interpretation
Gender			99.2%	The distribution of gender in the GSS sample is close to maximum variation. A 50%-50% split would result in an IQV of 100%.
Male	45.5%	1069		
Female	54.5%	1279		
Total	100.0%	2348		
Race identification			88.9%	With 2/3 of respondents being white-identified, this distribution is less variable than gender.
Whites	66.7%	1553		
Nonwhites	33.3%	776		
Total	100.0%	2328		
Race identification (detailed)			55.9%	With a more detailed classification, there's less variation in race identification than with just two categories. That's because no other category comes even close to being even with the percentage of whites.
White	66.7%	1553		
Black/African American	12.8%	299		
American Indian/Alaska Native	0.8%	18		
Asian Indian	0.8%	18		
Chinese	0.5%	12		
Filipino	0.6%	14		
Japanese	0.1%	3		
Korean	0.2%	4		
Vietnamese	0.3%	6		
Other Asian	0.8%	19		
Native Hawaiian	0.0%	0		
Guamanian/Chamorro	0.0%	0		
Samoan	0.0%	0		
Other Pacific Islander	0.1%	2		
Some other race	1.1%	25		
Hispanic	6.5%	152		
Multi-racial	8.8%	204		
Total	100.0%	2328		

Deeper Dive: Measures of Variability

Videos 14.6–14.9 provide more detail about measures of variability. These work best when viewed in the order shown.

What is Variability?

Range and Interquartile Range

Sum of Squares

Variance and Standard Deviation

Understanding Frequency Tables

You now have a good idea of how some basic tables and charts look in SPSS. If you try another data analysis program, such as SAS or Stata, the output you generate won't look exactly like what we showed you. The same is true of the tables and charts you'll encounter in the results sections of quantitative research papers. You'll need to take advanced coursework in statistical analysis to fully understand them, much less the terminology, symbols, and equations littered throughout the text. For now, our hope is that you learn enough of the fundamentals of quantitative data analysis to get the gist of what a researcher is trying to convey. We also hope to get you to the point that you can write a few sentences for each analysis that summarizes its conclusions concisely but with adequate detail for your audience. In this and subsequent sidebars, we'll discuss some basic principles for making sense of quantitative tables and writing an interpretation of their results. We'll start with frequency tables.



The example frequency table we examine in this section is from a journal article by Bob Lee, Yiwei Chen, and Lynne Hewitt (2011:1233), who surveyed older adults living at senior facilities in northwest Ohio about their use of information technology. *Matthias Zomer*, *via* *Pexels*

Government agencies and nonacademic research outlets often use frequency tables in their publications to provide descriptive statistics that describe a relevant population. Frequency tables are usually not the focus of academic papers—as we've pointed out, most academic sociologists seek to identify and explain relationships between two or more variables, generally speaking. However, you will often see frequency tables in a paper's methods section or the first part of its results section. The typical intent behind such a table is to offer a concise summary of the characteristics of the study's sample. **Figure 14.21** presents a table from an article by Bob Lee and his coauthors (2011:1233), which examined barriers encountered by older adults in using computer-mediated information technology.

Table 1
Respondents' social demographic profiles.

	N	%
Age		
50–64	103	44.2
65–74	79	35.1
75+	51	20.7
Level of education		
High school or less	41	17.3
Some college	67	28.3
Undergraduate	60	25.3
Graduate school	69	29.1
Residence		
Alone	71	30
Not alone	166	70
Income		
$10,000 or less	10	5.1
$10,001–20,000	21	10.7
$20,001–30,000	31	15.7
$30,001–40,000	21	10.7
$40,001–50,000	28	14.2
$50,001–60,000	24	12.2
$60,001+	62	31.5

Figure 14.21. Frequency Table from Lee et al. (2011). This frequency table from a journal article by Bob Lee, Yiwei Chen, and Lynne Hewitt (2011:1233) breaks down the distributions of four variables within the sample. *ScienceDirect*

When looking at any frequency table, your first task should be to determine what sample or population is being described, and what variable or variables are being featured. Let's take this approach for the example table. As a best practice, data tables will describe the sample briefly in a note that appears right under the data. This frequency table did not do that, meaning you would have to search the paper's methods section to learn that information. There, you would learn that this table is describing a nonprobability sample of 243 interviewees: older adults living at several senior facilities in northwest Ohio. The table title tells you that the distributions being displayed are for "social demographic" variables: age, education, income, and whether or not the person lives alone.

Now let's break down the table's data. The variables and their corresponding response options are labeled in the left column. The middle column (N) lists the number of cases that fall into the response category noted on the left. The right column (%) tells you the percentage of all cases that the count for that category represents. Note that when you add the percentages of all the response option rows for a single variable, they should equal 100 percent or something close. (In the latter case, the table note might warn readers: "Percentages may not sum to 100 due to rounding.")

Also note that this frequency table rounds percentages to a tenth of a percent. These authors are being particularly precise. For a sample this small—just a couple hundred respondents—most authors would choose to round the percentages to the nearest whole number. Either approach is generally fine, although you would want to be more precise if it is important to your research to distinguish between, say, 13.6 percent versus 14.4 percent (which, if rounded to whole numbers, would both be 14 percent).

Another best practice for data tables is to include a table note that lists the total sample size (if that number isn't already listed in a "total" row in the table body) and that also mentions how many missing values were recorded (which provides some indication about the quality of the survey). The example table doesn't do these things, but elsewhere in the article Lee and his collaborators (2011:1233) report they analyzed 243 surveys in total. The table shows 233 responses for the age question, so we can conclude that 10 respondents did not answer this question. That suggests our missing data for this question accounts for a little over 4 percent of the total sample—not too bad. While results like this would not make you doubt the overall quality of the survey, it helps to know where you might want to be cautious in your interpretations.

When you include a table in your own paper, you should break down its key findings in the main test. Don't just leave it to the reader to decipher the table. For interpreting tables, we recommend you follow the Generalization, Example, Exceptions (GEE) strategy described by Jane Miller (2015), which starts by generalizing the patterns, moves to providing one or more examples, and then concludes by mentioning any noteworthy exceptions. For univariate variables, the generalization step can be consolidated with the example step: you can flag the largest category or categories (generalization) and include their specific percentages (example), preferably mentioning them in descending order of size. You can also mention the percentages in parentheses after the category names. For instance, you might describe the educational attainment of the older adult sample described in **Figure 14.21** in the following way: "The sample was well-educated, with a plurality of respondents having a graduate degree (29.1 percent); another 25.3 percent had an undergraduate degree."

When you are interpreting a univariate table in prose, you do not have to list every single percentage that was listed in the table. The last percentage is obvious from context: if you list three out of four percentages, for instance, the percentage for the last category will just be the remaining amount for their sum to reach 100 percent. Furthermore, smaller categories may not be worth mentioning, especially if the variable has many different values.

The "exceptions" step of the GEE approach is often unnecessary when you are conducting a simple univariate analysis. That said, you might use the opportunity to flag a percentage that was surprisingly high or low. For instance, because the older adult sample was much more educated than the overall population of U.S. adults (who largely do not have college degrees), you might want to add another sentence to the earlier description: "Only 17.3 percent of respondents had a high school diploma or less education."

Quantifying and Comparing Life Outcomes: A Q&A with Elyas Bakhtiari



Elyas Bakhtiari is an assistant professor in the Department of Sociology at the College of William and Mary. His research examines how mortality rates and other health outcomes vary depending on factors like how countries form racial boundaries, how they incorporate immigrants into mainstream society, and how their social and political institutions shape the fundamental causes of health and illness. In other work, Bakhtiari draws on historical demographic records and machine learning to examine health outcomes among southern and eastern European immigrants in the early 1900s, Middle Eastern and North African populations after 2001, and other populations whose experiences of the U.S. racial and ethnic hierarchy was more complex than their official racial classification as "white" would indicate. His work has been published in such journals as *American Behavioral Scientist*, *Social Science and Medicine*, the *Journal of Health and Social Behavior*, and the *Journal of Racial and Ethnic Health Disparities*.

How did you get into sociology?

I started out as an engineering major and switched to sociology in part because I found it to be more intellectually stimulating and challenging in many ways. That's not to say undergraduate sociology courses were always harder than your average engineering course. But when dealing with technical problems, there are often clear and knowable answers that can be uncovered with enough analysis and background knowledge. With social problems, though, there often isn't a clear answer or single way to understand them—they result from thousands or millions of individuals consciously interacting in complex social systems.

I was drawn to sociology because understanding complex social phenomena not only seemed interesting but also important. Take something like climate change. We can solve some of the technical challenges by developing new technologies. But where we have really struggled is with collective action and behavior change, which are fundamentally social problems. It's the same thing with Covid-19. On the one hand, there's the technical achievement of developing numerous vaccines relatively quickly. On the other hand, there's the stubborn challenge of convincing enough people to take a vaccine that could save their lives. Sociology offers tools for understanding some of the major challenges of our time.

How do you use quantitative data in your own research?

I use quantitative data to study inequalities in health and mortality outcomes. Some of my work relies on large-scale surveys, in which people's answers to questions about their health, behaviors, and demographic characteristics have been quantified. I also rely on health records and vital statistics data, which includes birth and death records for entire populations. By analyzing patterns in how health and mortality outcomes vary, my

goal is to understand how people's health is affected by aspects of their social lives—particularly their experiences with racism, migration, and social inequality.

What advantages do you think quantitative research offers over other methodological approaches?

Quantitative research allows us to systematically compare outcomes. A lot of sociology focuses on how complex social processes can affect people's lives and life chances. If you're interested in a process, qualitative data can often provide a rich and nuanced perspective. But if you want to understand the outcomes of a process, you typically need some kind of quantified measure that allows for comparisons. In my work on health and mortality outcomes, I'm able to see how social processes have a profound impact on the quality and length of people's lives. Ideally, these two types of methods—qualitative and quantitative—work together to help us understand both the causes and consequences of social phenomena.

In your research, you have examined health and social inequalities in the United States and in Europe. Why do you think it's important to conduct comparative and cross-national research on inequalities?

Comparison can help us understand both the generalizability and variability of the social processes that affect population health—or any social phenomenon, really. Consider the relationship between health and socioeconomic status—that is, people's income and education levels. Cross-national comparisons have revealed that individuals with higher socioeconomic status have better health and longer lives across contexts, including in countries with universal health care or more equal income distributions. This is an important finding for advancing our theories about how socioeconomic status acts as a fundamental cause of health. However, we also see differences in the degree of the association across contexts. The relative health gaps between the high and low ends of the socioeconomic scale are larger in some countries than in others. This variability opens up a new set of research questions about what policies or social factors might shrink or widen relative health inequalities. Sociologists are often interested in how individuals are influenced by social contexts, and one of the clearest ways to see that is by comparing across different contexts.

What data and analytical challenges have you encountered when conducting comparative research?

It can sometimes be hard to find the data necessary to answer a comparative question. I am interested in how racism and ethnic discrimination affect health across national contexts, but some countries don't collect information on race or ethnicity in their censuses. Or even if data is available, each country has unique demographics, stratification systems, and constructed identities, which can make straightforward comparisons difficult. Luckily, there are some cross-national surveys that attempt to address these challenges by asking consistent questions about minority status across many countries. Although comparative questions are important, there is often a trade-off between breadth and depth when your data source spans a lot of different societies.

What advice do you have for sociology students who want to strengthen their quantitative skills?

The best way to strengthen quantitative skills is to never stop learning. I am constantly learning new quantitative skills during the course of my research. I try to choose my method based on what is best suited to answer the research question I'm interested in, and often that means learning a method I haven't used before. There are so many great resources for learning new methods now: coding tutorials, virtual classes, online forums, and even replication packages that allow you to see exactly how other researchers did their analysis. Although you can learn the foundational skills from your first statistics classes, the best way to build on that foundation is to put those skills into practice with an actual research project.

Key Takeaways

1. Use frequency tables and bar charts to display the frequency distributions of nominal-level and ordinal-level variables in terms of both counts and percentages.
2. It may be impractical to present scale-level data in a frequency table, but we can still calculate useful statistics about these variables, including the mean, median, and standard deviation.
3. Scale-level data can be visualized using a histogram or line graph, which will give you a sense of the shape of the distribution—whether it is left-skewed, symmetrical, or right-skewed.

14.3. Bivariate Data Analysis: Crosstabulations and Chi-Square



The "comfort theory" examined in this section refers to the idea that people in positions that allow for little satisfaction and fulfillment in everyday life turn to religion for solace. Based on that theory, one possible hypothesis is that being nonwhite makes it more likely people will seek out religious experiences. *Luis Quintero, via Pexels*

Learning Objectives

1. Understand how to conduct and interpret a crosstabulation analysis of two categorical variables.
2. Explain the logic behind null hypothesis testing and use it to interpret the results of a chi-square test.

So far, we have been conducting univariate analysis, generating statistics for one variable at a time. Now we will move

into bivariate analysis, which is about examining the relationship between two variables. Although we are far from being able to infer causality at this point, we should start thinking in terms of independent and dependent variables: specifically, what possible effect our **independent variable** (IV), also known as an *explanatory variable*, has on our **dependent variable** (DV), or *outcome variable*. (Recall that an independent variable is the variable we think may influence or explain another variable, the dependent variable.)

Let's say you have a general interest in learning about which types of people are more likely to be religious. You could develop your own survey to ask people about (1) their personal background and (2) their religious beliefs, and then use statistical analysis to see how these two sets of variables are related. Alternatively, you could find a quantitative dataset that already contains variables that cover this ground, and then start your analysis from there. As it turns out, many of the variables in the GSS relate to the importance of religion in respondents' lives (which would be your dependent variables for this particular analysis). But how do you choose independent variables from the many variables in the GSS that could give you a sense of a respondent's personal background? To put this question in another way, which demographic variables are likely to affect how devout a person is?

In situations like this, you should let theory guide you. Think about the literature you have read about the factors that influence people's religious beliefs. Which variables relating to people's identities have been brought up in past research? Use your creativity and logic to see how you might build on those theories and apply them to your own analysis.

Throughout the rest of this chapter, we will use an existing theory—"comfort theory"—to help us select independent variables and contextualize our bivariate analysis. According to this classic explanation for religious participation, the most devout individuals tend to be those whose "life situations most deprive them of satisfaction and fulfillment in the secular society" that surrounds them (Glock, Ringer, and Babbie 1967:107–8). In response, they turn to religion "for comfort and substitute rewards."

Working logically from this statement, we reason that members of marginalized social groups would be more likely to feel deprived of satisfaction and fulfillment. Therefore, they should also be more likely to turn to religion for comfort and sustenance.[1] To test this theory, we need to find some variables in the GSS data file that will allow us to identify more or less marginalized people. The following list includes the independent variables we plan to use, all of which we covered earlier in our univariate analyses. (Note that the corresponding variable name in the GSS data file is provided in parentheses, in case you want to follow along in SPSS.)

- **Age (*age*):** Elderly people have lower social status than young people.
- **Gender (*sex*):** Women have lower social status than men.[2]
- **Race identification (*racecen1*, *racecen2*, *racecen3*):** People identifying as members of racial or ethnic minority groups have lower social status than people identifying as members of the majority group.[3]
- **Educational level (*degree*, *educ*):** People with less formal education have lower social status than people with more

1. This analysis example is an extension of Earl Babbie's discussion in *The Practice of Social Research* (2016:49–50).

2. Although gender identification is a complex sociological concept, the GSS dataset provides only two binary gender categories, and we will use those two designations for the sake of simplicity. Since the GSS began in 1973, it has only explicitly captured the "sex" of respondents using its *sex* variable, but in more recent years, it has coded the genders of the respondent's household members with the variables *gender1*, *gender2*, *gender3*, and so on.

3. The GSS question asks, "What is your race? Indicate one or more races that you consider yourself to be." Respondents can choose up to three racial identities, with the first race mentioned in *racecen1*, the second in *racecen2*, and the third in *racecen3*.

education.[4]

To test the predictions of comfort theory, we also need to select dependent variables that plausibly capture what it means to turn to religion for "comfort and substitute rewards." There are numerous variables about religion to choose from in the GSS, but we have selected two that get at higher and lower levels of participation in religious activities. Since these dependent variables are a bit less generic than our independent variables, this time, we have included the exact text of our questions—which is crucial to refer to whenever you interpret the results of an analysis.

- **Frequency of attending religious services (*attend*):** "How often do you attend religious services?" This question gets at the extent to which a person is willing to publicly express their commitment to their religion.
- **Frequency of prayer (*pray*):** "About how often do you pray?" This question gets at the extent to which a person expresses their religious devotion in private.

Let's pause here to acknowledge the **deductive approach** we are pursuing with this bivariate analysis. First, we drew upon previous research to identify a relevant theory—that is, the comfort theory of participation in religion. We then used that theory to help us settle on sets of independent and dependent variables. We selected four independent variables that serve as reasonable indicators of whether our respondents are members of marginalized groups. And we selected two dependent variables that serve as reasonable indicators of the importance of religion in their lives.

Next, we will generate **hypotheses** from our theory—specific and testable propositions about what the empirical results of a study will be if the underlying theory is correct. As we discussed in **Chapter 5: Research Design**, our hypotheses should clearly and unambiguously state what we expect to find when we do the data analysis. Here are some hypotheses about how our independent variables are expected to influence our dependent variables if comfort theory does indeed hold up in this context:

1. The higher a respondent's age is, the higher their frequency of attending religious services will be.
2. Women will have a higher frequency of attending religious services than men will.
3. Nonwhites will have a higher frequency of attending religious services than whites will.
4. The lower the respondent's educational level is, the higher their frequency of attending religious services will be.

We can create another set of four hypotheses by swapping out the variable for frequency of attending religious services with the second dependent variable, frequency of prayer—for instance, "Women will have a higher frequency of prayer than men will." Note as well that all these hypotheses can be stated in the reverse: "The lower the respondent's age is, the lower their frequency of attending religious services will be."

Now it's time to analyze our data and test our hypotheses. Before we do that, however, we should point out one more thing. While quantitative researchers frequently take the deductive approach to their data analysis that we have just outlined, an **inductive approach** is also an option. Working inductively, for instance, we could explore our data to examine a wide variety of variables that might influence a person's degree of religious faith. We could develop a theory based on the results of this analysis, which might extend comfort theory or cover entirely different ground. (For more on the inductive approach, refer to **Chapter 4: Research Questions**.)

4. The GSS has two measures of a respondent's educational attainment: *degree* and *educ*. *Degree* is an ordinal-level variable that indicates the highest educational diploma or degree that the respondent has completed: less than high school, high school, associate/junior college, bachelor's, and graduate. *Educ* is a scale-level variable that tells us the highest number of years of formal education completed by the respondent: 0, 1, 2, 3, and so on, up to a high of 20.

Bivariate Crosstabulation: Analyzing Two Categorical Variables

The type of bivariate analysis we can conduct depends on the level of measurement of the independent and dependent variables we are analyzing. *When all our variables are categorical variables (measured at the nominal or ordinal level)*, we can use **crosstabulation analysis**. This analytical procedure creates a table showing how the categories of multiple variables intersect. We could throw any two variables into this table just to see if a relationship exists between them, but typically we have an independent variable and a dependent variable in mind, with the goal of examining the possible effect of the former on the latter. In this section, we will use *crosstabulation tables* to evaluate the hypotheses we derived from comfort theory, examining how our independent variables of age, race, gender, and education relate to our dependent variables of frequency of attending religious services and frequency of prayer. These tables—usually just called **crosstabs—**also go by the names *contingency tables* and *two-way tables* (the latter when they encompass just two variables).

Crosstabulation analysis is a good choice for examining relationships between an independent variable and dependent variable when two conditions are met: (1) as we already mentioned, both variables should be measured at the nominal or ordinal level; and (2) both variables should have a narrow range of possible values (i.e., few response options for each survey question)—typically, *between two and five values each*. Condition 2 means that if your categorical variable happens to have numerous response categories, you'll want to *collapse* those categories to only a few. For instance, if you were collapsing the many categories included in the GSS for its *racecen1* variable (which includes all of the race options on the U.S. Census form), you could keep the "White" and "Black" categories as they are, but combine (collapse) all the other race categories into an "Other" category. (In fact, those are the three categories in the original GSS *race* variable, which has stayed more or less consistent since the survey began in 1972.)

Condition 1—that your crosstab's variables should be nominal or ordinal level—suggest that you can't include a scale-level variable like the GSS age variable in its raw form. You can, however, easily convert a scale-level variable into an ordinal-level variable that *can* be analyzed with a crosstab. To do so, you will need to collapse the many response categories of your scale-level variable to only a few. You can think of these new, consolidated categories as buckets in which you are tossing a range of values (in fact, one term for them is **bins**). For example, we might sort the values of our scale-level variable *age* into just four broad bins: 18 to 34, 35 to 49, 50 to 64, and 65 and over.

Collapsing and combining values of a variable is a common method of *recoding*. When you recode a variable, you change the original values to make them easier to analyze. Once we recode the *age* in the way we have described, we can still rank our respondents by their ages from low to high (or high to low) because our new variable is ordinal-level. But we have lost the more precise measurement of exact years. (This is one reason that you generally do not want to replace an existing variable with a recoded one; instead, create a new variable and save the old one in case you realize later that you need greater precision.)

Let's go ahead and generate a crosstab showing the relationship between gender (IV) and frequency of prayer (DV). Gender is nominal level and has two response categories. Frequency of prayer is ordinal level and has six categories—not ideal, but okay for this example. Will the analysis support our hypothesis that women, the historically marginalized group, will have higher frequencies of prayer than men?

Figure 14.22 depicts the crosstab we generated in SPSS. For this example, we chose to put our dependent variable in the rows, and our independent variable in the columns. Structuring your table in this way is probably the approach you want to use when you're first learning to do crosstabulation analysis, but we'll show you later what happens when you swap what variables go in the table's rows and columns.

How often Does R pray * Gender Crosstabulation

Count

		Gender		
		Male	Female	Total
How often Does R pray	SEVERAL TIMES A DAY	222	458	680
	ONCE A DAY	270	379	649
	SEVERAL TIMES A WEEK	112	119	231
	ONCE A WEEK	82	70	152
	LT ONCE A WEEK	137	113	250
	NEVER	240	130	370
Total		1063	1269	2332

Figure 14.22. Crosstabulation of Frequency of Prayer (DV) by Gender (IV) with Counts. When we put two variables in a *crosstabulation table* (also called a *two-way table*), we have the option of putting the presumed independent variable in the columns or rows. In this example, we've put the independent variable, gender, in the columns, so we see "Male" and "Female" as the headings of our columns.

There's a lot going on in this crosstab. Let's break down its different elements:

Title: For crosstabs, SPSS automatically generates a title. The default title is in the format "[*Row variable*] * [*Column variable*] Crosstabulation." The asterisk stands for "by"—as in "length by width" (rows by columns). You can edit the table title in SPSS (or whatever program you are using) to provide a more descriptive title, perhaps something like "How often respondent prays by their gender" or "Frequency of prayer by gender."

Red boxes: Twelve table cells in which the two gender categories intersect with the six prayer categories. Each cell includes the count of people who fit the conditions in that column and that row. Let's take the two cells in the **green box** as examples. The left cell indicates that 222 men said they pray several times a day. The right cell indicates that 458 women said they pray several times a day.

Purple box: Totals for the gender categories. The left cell indicates that 1,063 men answered the prayer question (with any of the six response options), and the right cell indicates that 1,269 women did so.

Blue box: Totals for the prayer categories. The top cell indicates that 680 respondents (both men and women) said they pray several times a day, and the bottom cell indicates that 370 respondents said they never pray.

Orange box: Total number of responses to the prayer question, counting both men and women. This is the sum of the counts in all 12 intersecting cells (the red boxes).

Based on the cells in the green box, we know that more women than men in our sample pray several times a day. Does this mean that in the U.S. adult population (the sample's target population) women are more likely than men to pray several times a day? Not necessarily. The cells in the purple box tell us that about 200 more women than men answered the prayer questions. The greater number of women than men in that "Several times a day" category might just be because there are more women than men in our sample. We need to take into consideration the size of each of these groups when considering whether any patterns we observe in the cell counts are meaningful.

Luckily, this has an easy fix. We simply add **column percentages** to the table—that is, percentaging each cell by using its count as the numerator and the column total as the denominator. Since we put the independent variable, the respondent's gender, in the crosstab's columns, we use the totals at the bottom of the "Male" and "Female" columns—1,063 and 1,269, respectively—to calculate the cell percentages. (Each column total serves as the **percentage base**—the denominator when we calculate the percentage for each cell in that column.) **Figure 14.23** presents the revised crosstab percentaged by columns.

How often Does R pray * Gender Crosstabulation

			Gender		
			Male	Female	Total
How often Does R pray	SEVERAL TIMES A DAY	Count	222	458	680
		% within Gender	20.9%	36.1%	29.2%
	ONCE A DAY	Count	270	379	649
		% within Gender	25.4%	29.9%	27.8%
	SEVERAL TIMES A WEEK	Count	112	119	231
		% within Gender	10.5%	9.4%	9.9%
	ONCE A WEEK	Count	82	70	152
		% within Gender	7.7%	5.5%	6.5%
	LT ONCE A WEEK	Count	137	113	250
		% within Gender	12.9%	8.9%	10.7%
	NEVER	Count	240	130	370
		% within Gender	22.6%	10.2%	15.9%
Total		Count	1063	1269	2332
		% within Gender	100.0%	100.0%	100.0%

Figure 14.23. Crosstabulation of Frequency of Prayer by Gender with Column Percentages. If we put the independent variable (here, gender) in the columns of our crosstab, it is best to percentage by columns, so that we can easily see how the independent variable may be influencing answers to the dependent variable.

By adding percentages, we are finally able to compare men and women by how frequently they pray in an apples-to-apples fashion. As shown in the blue ovals, the percentages for the two groups sum to 100 percent. That's what makes it possible to compare them even though we have 200 more women than men in our sample: the percentages show us what the distribution would be like if each group had the same number of respondents. (It's as if they each had 100 respondents!)

With the percentages in place, we can see how men and women differ across the categories of the dependent variable, frequency of prayer. But which of the six response categories should we focus on? It depends on our hypothesis. In this example, our hypothesis is that women will pray more frequently than men do, so we might want to focus on the category indicating the highest frequency of prayer, "Several times a day." Did more women pick this option? Yes, they did (see the cells in the red box), which is in line with our theory. We might also want to examine the category for the lowest frequency of prayer ("Never"). Did more men pick this option? Yes, they did (see the cells in the green box), providing more support for our hypothesis.

Note that having many response categories complicates the sort of comparison we've been talking about. For instance, what if fairly similar proportions of women and men said they prayed several times a day, but many more women than men prayed once a day? Is that evidence in support of our hypothesis? Perhaps, and in this case, we might want to collapse our response options to fewer categories so that we are better able to see overall patterns. After all, when a survey question has numerous response options, it is common for an extreme option to attract very few respondents—making it hard to know if any differences across groups in choosing that extreme option really tell us something meaningful. In this case, we might combine the extreme response category with a closely related category. For example, if few cases were included in the "Several times a day" category, we could merge it with "Once a day" and then examine differences between men and women in the combined "At least once a day" category. Likewise, if we were dealing with a Likert-scale variable, we might want to combine the "Strongly agree" and "Agree" response options into a single value (perhaps just called "Agree"), and do the same for "Strongly disagree" and "Disagree." Then we'd be accounting for how the independent variable might affect overall agreement or disagreement with the statement in question, rather than focusing narrowly on the IV's influence on a single (extreme) response option like "Strongly agree."

The crosstabs we have shown so far have displayed the independent variable as the *column variable*. Does a crosstab always have to be done this way? No, and sometimes researchers have good reasons to put the independent variable in the rows—for instance, when the independent variable has many categories, or when they are analyzing multiple independent variables (with crosstabs stacked on top of each other, essentially). That's because it is generally easier to create tables with numerous rows than numerous columns—for instance, if you were writing a paper, you might have to change the page orientation to landscape mode to accommodate lots of columns.

If, however, you choose to make the independent variable the *row variable* in your crosstab, you must have your data analysis program calculate **row percentages**. Figure 14.24 depicts a crosstab with the independent variable listed this time in the rows. Note how the title for the crosstab has changed from what it was in the crosstab percentaged by columns, given that the default SPSS title takes the form "[*Row variable*] * [*Column variable*] Crosstabulation."

With percentaging by rows, each row category of the independent variable now adds up to 100 percent (as shown in the blue ovals). We can compare the two IV categories, men and women, across rows within one column of the DV at a time (as shown in the red box for "Several times a day" and in the green box for "Never"). The percentages given in this table are identical to those for the same bivariate analysis in **Figure 14.23**—even though we transposed the columns and rows—because we also changed the percentaging.

Gender * How often Does R pray Crosstabulation

			How often Does R pray						
			SEVERAL TIMES A DAY	ONCE A DAY	SEVERAL TIMES A WEEK	ONCE A WEEK	LT ONCE A WEEK	NEVER	Total
Gender	Male	Count	222	270	112	82	137	240	1063
		% within Gender	20.9%	25.4%	10.5%	7.7%	12.9%	22.6%	100.0%
	Female	Count	458	379	119	70	113	130	1269
		% within Gender	36.1%	29.9%	9.4%	5.5%	8.9%	10.2%	100.0%
Total		Count	680	649	231	152	250	370	2332
		% within Gender	29.2%	27.8%	9.9%	6.5%	10.7%	15.9%	100.0%

Figure 14.24. Crosstabulation of Frequency of Prayer by Gender with Row Percentages. If we place our independent variable in the rows of the crosstab, we will want to percentage by rows instead of columns—as is the case in this table, where the percentages in each row sum to 100 percent. Note that the counts and percentages in the cells are exactly the same as they were in the crosstab in Figure 14.23 that was percentaged by columns and placed the independent variable in the columns.

We've given you a quick sense of how researchers go about comparing percentages across groups within a crosstab. Let's go over how to make sense of crosstabulation results in greater detail. When conducting this analysis, you should ask yourself the following questions (what we'll call our *crosstabulation analysis questions*):

1. **Does a relationship exist between the independent and dependent variables?** Take a look at the percentages in the cells. Do they differ across categories of the independent variable? For instance, do respondents answer the survey question of interest (your dependent variable) differently depending on their characteristics (your independent variable)?

    1. If **no**, then stop. No relationship exists between the independent and dependent variables.
    2. If **yes**, then a relationship may exist between the independent and dependent variables. Continue with Questions 2 and 3.
2. **Is there a pattern to the relationship between the independent and dependent variables?** How exactly do respondents answer the question of interest (your dependent variable) differently depending on their characteristics (your independent variable)? For instance, can we identify a general pattern of higher support for a particular kind of response among respondents from one group?
3. **If both variables are ordinal-level, is there a direction to the relationship (positive or negative)?** We will discuss the direction of relationships later in the chapter.
4. **What is the strength of the relationship?** How closely are the two variables associated with each other? Is the relationship strong or weak?

There are many ways to determine the strength of a relationship between categorical variables. We'll cover a very simple approach: the **percentage-point difference method**. Start by deciding which category of the dependent variable is most theoretically important to your analysis. For instance, in the earlier crosstab depicted in **Figure 14.23**, which had six categories for the DV, you might focus on the "Several times a day" category, the highest possible frequency of prayer. Out of all the DV categories, the percentages of responses that fall into this category (and perhaps the "Never" category as well) should provide the most useful information in helping you determine which group is more likely to turn to religion for comfort—the theory we want to test.

Now that you've settled on which DV category to focus on, take a look at the cells associated with that category. If the crosstab's independent variable has only two values (e.g., male and female), subtract the smaller percentage from the larger percentage; if the IV has three or more values (e.g., ages 18–39, 40–64, and 65 and older), subtract the smallest percentage from the largest percentage. *The greater the percentage-point difference, the stronger the relationship between the IV and DV.* Roughly speaking, a difference of 5 percentage points or more could signal a relationship worthy of further investigation. (Note that we are speaking here of the percentage-*point* difference, and not a *percentage* change.[5])

What do the various outcomes described by our crosstabulation analysis questions look like in an actual table? **Figure 14.25** shows a hypothetical example with four possible results for the crosstabulation of gender (IV) with frequency of

5. The *percentage-point difference* is calculated by subtracting the smaller percentage from the larger percentage, and this is the comparison you want to make between two percentages in a crosstab. If you were looking at a change over time, however, you could measure that change either as (a) a percentage-point difference or as (b) a *percentage change* (calculated as the percentage-point difference between the earlier and later values divided by the earlier value). For instance, if the percentage of women who pray several times a day fell from 36 percent in 2018 to 34 percent in 2022, we could say that the rate of women praying this frequently fell by 2 percentage points, or that it declined by 6 percent (a 2 percentage-point difference divided by the earlier value of 36 percent); you could not say it declined by 2 percent.

prayer (DV). The percentages have been chosen to illustrate different scenarios. Remember that gender is measured at the nominal level, and frequency of prayer at the ordinal level. Note as well that we have collapsed the six categories of the frequency of prayer variable to just two categories so that we can more easily see how men and women differ overall in how often they pray.

A

	Gender		
Frequency of prayer	**Male %**	**Female %**	**Totals**
Once a week or less	55% ⟷ 20%		36% (160)
Several times a week or more	45% ⟷ 80%		64% (290)
Totals	100% (200)	100% (250)	100% (450)

B

	Gender		
Frequency of prayer	**Male %**	**Female %**	**Totals**
Once a week or less	36% ⟷ 36%		36% (162)
Several times a week or more	64% ⟷ 64%		64% (288)
Totals	100% (200)	100% (250)	100% (450)

C

	Gender		
Frequency of prayer	**Male %**	**Female %**	**Totals**
Once a week or less	0% ⟷ 100%		56% (250)
Several times a week or more	100% ⟷ 0%		44% (200)
Totals	100% (200)	100% (250)	100% (450)

D

	Gender		
Frequency of prayer	**Male %**	**Female %**	**Totals**
Once a week or less	56% ⟷ 48%		52% (232)
Several times a week or more	44% ⟷ 52%		48% (218)
Totals	100% (200)	100% (250)	100% (450)

Figure 14.25. Possible Relationships between Variables in a Crosstab. The four crosstabulation tables in this diagram represent four (of many) possible outcomes of a crosstabulation analysis comparing gender and frequency of prayer. In Table A, women pray much more often than men do. In Table B, men and women pray just as often. In Table C, there is a "perfect" relationship between frequency of prayer and gender: all women pray frequently, and all men pray infrequently (as defined by the two collapsed categories of our recoded frequency of prayer variable). Finally, in Table D, men pray more often than women do.

Let's apply our crosstabulation analysis questions to these four tables:

1. **Does a relationship exist between the independent and dependent variables?** Tables A, C, and D show such a relationship. In each case, the percentages of men and women who pray several times a week or more are different from one another. Table B is the only crosstab that shows no discernible relationship—that is, no difference between men and women in how often they pray.
2. **Is there a pattern to the relationship between the independent and dependent variables?** In Tables A and D, women are more likely than men to pray frequently. In Table C, men are more likely than women to pray frequently.
3. **If both variables are ordinal-level, is there a direction to the relationship (positive or negative)?** We can't evaluate direction, since gender is a nominal-level variable.
4. **What is the strength of the relationship?** We'll use the percentage-point difference method to evaluate the strength of these four bivariate relationships. Because we have collapsed the six categories of the dependent variable to just two—"Once a week or less" and "Several times a week or more"—deciding which DV category to focus on is easy. Either category works fine because the percentage for one of the DV categories will be the inverse

of the other. (To use Table A as an example, if 55 percent of men pray once a week or less, then 45 percent must pray several times a week or more, given that those are the only two DV categories.) Given that knowing the percentage for one DV category tells you the percentage for the other, analyzing either category gives you the same information, and either category can be used to calculate the percentage-point difference. Let's look at the DV category "Several times a week or more." In Table A, there is a 35 percentage-point difference between the two genders (80 − 45 = 35), in Table B a zero percentage-point difference (64 − 64 = 0), in Table C a 100 percentage-point difference (100 − 0 = 100), and in Table D an 8 percentage-point difference (52 − 44 = 8). Based on these calculations, we conclude that the strongest relationship is found in Table C, followed by A, and then D. In Table B, there is a 0 percentage-point difference, indicating no relationship exists between the IV and DV.

Realistically, outcomes like those given in Table B (where the percentages are the same across genders) and Table C (where one gender falls entirely in the category of frequent prayer) rarely, if ever, occur in sociology. *Some* sort of relationship—some sort of difference across values of the IV—is inevitably detectable in the sorts of data we sociologists analyze, given the myriad factors that shape outcomes in the social world. That said, generating the relevant crosstab can usually give us an intuitive sense of how strong a relationship is (including whether it plausibly exists, given the percentage-point differences we actually see). And as we will discuss later, we can also calculate *inferential statistics* to determine if any differences we do see are likely to be due to a real relationship between the variables in our target population, or if they are quite possibly due to a fluke sample we happened to collect.

At this point, though, focus on figuring out exactly what your crosstab is telling you. Write up a brief interpretation of it that you can insert into the main text of your paper, which should always discuss the key takeaways of any table or chart you include. When summarizing your results, refer to your answers to the crosstabulation analysis questions, in addition to keeping the following things in mind:

- Before describing any results, state the hypothesis that you are testing with your crosstabulation analysis. (See **Chapter 5: Research Design** for more advice on constructing hypotheses.)
- To reiterate, a write-up should not mention every percentage in your table. That's what the table is for! Describe the results that are most important to your research question. For instance, "Several times a day" and "Never" are probably the most theoretically relevant categories in the crosstab of frequency of prayer by gender depicted in **Figure 14.23**, since they measure the two extremes of greatest and least frequency of prayer. You might not need to mention the other categories, or you might want to consolidate them and just report the collapsed categories.
- In your interpretation of the findings, simplify percentages by rounding to the nearest whole percent. Distinctions on the basis of tenths of a percent are usually too small to make any meaningful difference. If you like, you can present the percentages in the table with a single decimal place (in most cases, just one is more than enough).

Let's revisit the actual GSS data on gender and frequency of prayer. Using the crosstab in **Figure 14.23** with column percentages, let's review our list of crosstab questions to evaluate the relationship between these two variables:

1. **Does a relationship exist between the independent and dependent variables?** Yes. Men and women show different percentages within categories of the dependent variable, frequency of prayer.
2. **Is there a pattern to the relationship between the independent and dependent variables?** Yes. On the high end of frequency of prayer, 36 percent of women pray several times a day, whereas 21 percent of men do so. On the other end of the distribution, 10 percent of women never pray, whereas 23 percent of men never do so. Compared to men, then, women are more likely to pray several times a day and less likely to never pray.
3. **If both variables are ordinal-level, is there a direction to the relationship (positive or negative)?** We can't evaluate direction, since gender is a nominal-level variable in this table.
4. **What is the strength of the relationship?** There is a 15 percentage-point difference between the two genders within the "Several times a day" category and a 13 percentage-point difference within the "Never" category.

Now that we have interpreted the results of our crosstab, how would we go about writing them up in the findings section of a paper? To summarize the results of crosstabs and other bivariate analyses, we recommend you follow Miller's (2015) GEE approach described earlier. First, summarize any overall pattern you found. Then provide one or more example statistics that illustrate that pattern. Finally, mention exceptions to the observed pattern (if any). Take a moment now to write a few sentences describing the data given in **Figure 14.23**. Then compare your write-up to the wording we suggest in the answers to the exercises at the end of this section.

Up to this point, we have skipped Question 3, which is about determining the direction of a bivariate relationship. That's because one of our two variables (gender) was at the nominal level, which means there cannot logically be a direction to the relationship between that variable and another variable. If, however, both of our variables are measured at the ordinal level, we can determine the direction of their relationship. Here, we are looking for one of two patterns:

1. If high values of the IV tend to occur along with high values of the DV, we say the relationship is a **positive relationship**.
2. If high values of the IV tend to occur along with low values of the DV, we say the relationship is a **negative relationship** (also known as an *inverse relationship*).

Figure 14.26 contains four crosstabs with hypothetical examples of some of the possible relationships between age and frequency of attending religious services. We have measured both of these variables at the ordinal level, which allows us to conduct a crosstabulation analysis that describes the direction of their relationship.[6]

6. Both of these variables could be measured in a scalar form. For example, we could measure the age variable as the respondent's age in years and the frequency of attending religious services variable as the number of times the respondent attended religious services last month. Collapsing these variables into ordinal-level measures with much fewer categories allows us to use them in a crosstabulation analysis, but it also means that we lose some information—specifically, the ability to tell exact distances between values.

A

Attendance	Age		
	Young	Middle-Aged	Senior
Weekly	0%	0%	100%
Occasional	0%	100%	0%
Never	100%	0%	0%
Total	100%	100%	100%

B

Attendance	Age		
	Young	Middle-Aged	Senior
Weekly	100%	0%	0%
Occasional	0%	100%	0%
Never	0%	0%	100%
Total	100%	100%	100%

C

Attendance	Age		
	Young	Middle-Aged	Senior
Weekly	22%	35%	56%
Occasional	28%	24%	17%
Never	50%	41%	27%
Total	100%	100%	100%

D

Attendance	Age		
	Young	Middle-Aged	Senior
Weekly	65%	33%	20%
Occasional	24%	47%	31%
Never	11%	20%	49%
Total	100%	100%	100%

Figure 14.26. Possible Directions of Relationships in a Crosstabulation. The four crosstabulation tables in this diagram represent four (of many) possible outcomes of a crosstabulation analysis comparing age and frequency of attending religious services. In Tables A and B, there is a "perfect" relationship between the two variables, with a person's age (as defined by the listed age categories) fully explaining where that person falls in terms of attending religious services (as defined by the listed frequency categories). However, the perfect relationship in Table A is perfectly positive—as age increases, attendance at religious services rises—whereas in Table B, it is perfectly negative—as age increases, attendance declines. Tables C and D show weaker positive and inverse relationships, respectively, between the same two variables.

Let's compare the different relationships depicted in **Figure 14.26** by asking some questions about them:

1. **Which of these tables displays a positive relationship between the independent variable (age) and the dependent variable (frequency of attending religious services)?** Tables A and C show a positive relationship between age and frequency of attending religious services. As age increases, frequency of attending religious services increases. Said another way: the higher the age, the higher the likelihood of frequent attendance. The reverse is also true: the lower the age, the lower the likelihood of frequent attendance. In fact, in Table A, the relationship is a "perfect" positive relationship: all seniors have high (weekly) attendance, and all young people have low (never) attendance, with middle-aged people falling in between. (Again, you will rarely see such extreme results in real life.)
2. **Which of these tables displays a negative relationship between the independent variable (age) and the dependent variable (frequency of attending religious services)?** Tables B and D display a negative relationship between age and frequency of attending religious services. As age increases, frequency of attending religious services decreases. The higher the age, the lower the likelihood of frequent attendance. The reverse is also true: the lower the age, the higher the likelihood of frequent attendance. In fact, in Table B, the relationship is a "perfect" negative/inverse relationship. All seniors have low (never) attendance, and all young people have high (weekly) attendance, with middle-aged individuals falling in between.

Now that we have covered the basic principles of bivariate crosstabulation analysis, we can use these techniques to evaluate all the comfort theory hypotheses at the same time. **Table 14.5** and **Table 14.6** display the results of crosstabs for each of the two dependent variables, frequency of prayer and frequency of attending religious services. To keep you on your toes, we've put the four independent variables in the rows of each table and percentaged them by rows—a common

thing to do, as we've noted, when you are analyzing many IVs at the same time. Furthermore, in **Table 14.5**, we have used the recoded frequency of prayer variable that has just two values.

As you did for the crosstab of frequency of prayer (with six categories) and gender in **Figure 14.23**, write down a few sentences describing the relationships described in the two tables. Then compare them with the answers to Exercises 1 and 2 at the end of this section (note that these are suggestions rather than answers, given that there are many perfectly acceptable ways to write up your results). Remember to use the crosstabulation analysis questions we posed earlier to determine the pattern and strength of each relationship (and also its direction if both variables are ordinal-level). Then use the GEE approach to concisely describe those results.

Table 14.5. Crosstabulations of Frequency of Prayer by "Comfort Theory" Independent Variables

		Frequency of Prayer		
		Once a week or less	**Several times a week or more**	***n***
Age	**18–39**	46.6%	53.4%	932
	40–64	24.8%	75.2%	971
	65 and older	23.2%	76.8%	425
Gender	**Male**	43.2%	56.8%	1064
	Female	24.7%	75.3%	1270
Racial identity	**White**	36.3%	63.7%	1547
	Nonwhite	26.7%	73.3%	770
Have a college degree?	**No**	32.5%	67.5%	1619
	Yes	34.5%	65.5%	716

Table 14.6. Crosstabulations of Religious Service Attendance by "Comfort Theory" Independent Variables

		Frequency of Attendance at Religious Services			
		Once a year or less often	**One to three times a month**	**More than three times a month**	***n***
Age	**18–39**	56.9%	15.5%	27.6%	934
	40–64	43.3%	18.8%	37.9%	967
	65 and older	41.2%	12.7%	46.1%	424
Gender	**Male**	53.4%	16.0%	30.7%	1064
	Female	44.0%	16.8%	39.3%	1267
Racial identity	**White**	50.9%	16.5%	32.7%	1547
	Nonwhite	42.9%	16.3%	40.8%	767
Have a college degree?	**No**	51.9%	14.8%	33.3%	1615
	Yes	40.1%	20.0%	40.0%	716

General Tips for Reading Data Tables

Now that we have covered both univariate and bivariate tables, the following general tips should help you evaluate all kinds of data tables that appear in academic publications:

1. When reading a paper, at least skim over its methods section before you attempt to understand one of its data tables.
2. Start by carefully reading the table title first. If it is a well-written title, it will give you hints about the nature of the variables or types of analyses represented in the table. A table title sometimes includes background information, such as the dataset name and year and the target population, all of which can be concisely packed into a single line—for example, "Belief that abortion for any reason should be legal among U.S. adults (GSS, 2018)." That said, these details are frequently provided in the *table note* instead (see item 4). For bivariate analyses, a common approach is to list the dependent variable first, followed by the words "by" or "as a function of" and then the independent variable: for instance, "Frequency of attending religious services by respondent's age" or "Frequency of prayer as a function of educational attainment."
3. Read column and row labels to further understand the table contents, paying particular attention to any

variables listed. Attempt to piece together the role of each variable in the analysis, including whether the researchers are treating it as an independent or dependent variables (or even as a moderating or mediating variable, as we described in **Chapter 3: The Role of Theory in Research**).

4. Read the **table note**—the text right underneath the table. If such details were not provided in the title, the table note should prominently indicate the dataset and year (e.g., "Source: General Social Survey, 2018"). It may also convey other important information about the data, such as how key variables were measured, what target population the sample was drawn from, how many cases were in the sample, and how many missing values were excluded from the analysis. For bivariate and other multivariate analyses, the table note should provide a legend (or explanatory text) that describes what thresholds for statistical significance were used for any hypothesis testing (e.g., * $p < .05$; ** $p < .01$; *** $p < .001$). In accordance with the legend in the table note, any statistically significant results should be denoted with asterisks or other symbols within the table itself.
5. Consider the data analysis technique or techniques represented in the table (e.g., crosstabulation analysis, correlation matrix, regression analysis). The specific methods that researchers used may be described in the table title or table note. Oftentimes, a table that appears complex can be broken down into a series of separate analyses.
6. Relate what you see in the data table to the purpose of the study and the concepts that guide it.

Probably item 6 is the most important tip. If a table does not contribute to describing or understanding aspects of the social phenomenon being studied, it's worthless, no matter how fancy the table structure or sophisticated the data analysis techniques. The links between theory and data must be strong to provide any truly useful information.

Inferential Statistics: Chi-Square Analysis



Once we have analyzed any patterns in our data, we need to determine whether they can be generalized to our target population. Assuming we have collected an unbiased probability sample to begin with, the danger is that we may be analyzing a sample that, by chance, poorly represents the population of interest. Much like picking a wildly different specimen out of an assortment of fruits could give us a skewed perspective on that kind of fruit, a fluke sample may be showing us patterns that don't exist in the actual population of interest. *Nathan J. Hilton, via Pexels*

So far, we have described patterns within our sample. But in deductive quantitative analysis, we usually want to generalize any results we derive from our sample to our target population. That's the only way we can test the theories we want to test, all of which must apply to the larger population to be valid.

Can we generalize the results we have obtained through our crosstabulation analysis of GSS data? To answer this question, we can generate *inferential statistics* that specifically tell us if it is acceptable to use our sample results to make general statements about what's happening in the target population. An important condition has to be met, however, before we can proceed: our sample must have been gathered using probability sampling (as described in **Chapter 6: Sampling**). If that is not the case, then inferential statistics are not appropriate, and we cannot generalize our sample's results—no matter how large our sample is, or how great our survey questions were. Fortunately, the GSS sample is a probability sample collected from a target population of American adults, so we can calculate inferential statistics to evaluate whether each of our crosstab results can be applied to the U.S. adult population.

What do inferential statistics tell us? As a thought exercise, imagine that there really is no relationship between gender and frequency of prayer in the U.S. adult population. Men and women pray with the same frequency. If we selected many, many random samples of, say, 2,000 adults from the U.S. population, most of the samples would also show no relationship. No relationship actually exists in the population, and if our selection of respondents is truly random, most of the samples we collect will reflect that fact. But the GSS sample happens to show a relationship between gender and frequency of prayer. Clearly, the sampling was off this time—a rare occurrence with random sampling—because it doesn't reflect the true situation in the population of interest.

This hypothetical scenario is a risk any time that we conduct a bivariate analysis and observe a relationship between two variables. We need to ask if we are sure there's a real relationship in the target population, or if—as in the scenario we just proposed—we happen to have a weird sample that doesn't represent the target population well. Such an outcome is unlikely if we truly did secure a random sample from our population of interest, but it is still possible, and it raises the risk that we will erroneously conclude based on our fluke sample that a relationship exists in the population.

We've just described the basic framework of **null hypothesis testing** (for a more detailed discussion of this critical concept, see the sidebar **Understanding Null Hypothesis Testing**). A null hypothesis test is a formal procedure for evaluating the hypothesis that there is no relationship between two (or more) variables in the target population. For crosstabulations, the most common null hypothesis test is called the **chi-square test** (pronounced *KAI-square*). For our analysis of the relationship between gender and frequency of prayer, the null hypothesis for our chi-square test would be that there is no such relationship in the larger population. If we can reject this idea of no relationship, our argument that the two variables are related in the actual U.S. adult population is strengthened.

Note the careful wording of this statement. With bivariate crosstabulations, we are determining whether or not there is a *relationship* between the two variables. We are not able to determine if one variable *causes* the other. Unless we use more advanced methods, we can only detect **correlation** (also known as *association*)—that is, whether a change in the independent variable is associated with a change in the dependent variable. We can say the two variables are "related," "associated," "linked," or "correlated," but we cannot say that a change in one causes a "change" in another. As you may recall from our discussions of causality in **Chapter 3: The Role of Theory in Research** and **Chapter 12: Experiments**, correlation is a necessary condition for a cause-effect relationship to exist, but by itself it is insufficient for establishing

causation. Among other things, we also need to rule out alternative explanations—perhaps by using multivariate analysis techniques to account for those possibilities, as we will discuss later.

One common standard for testing a null hypothesis is this: reject the null hypothesis if the chance of being wrong is less than 5 percent (.05). This is the conventional threshold for what we call **statistical significance**—having a high degree of confidence that the relationship we have observed in the data is not due to a sample that by chance does not adequately reflect the characteristics of the target population. Using this threshold, let's evaluate the hypotheses we've derived from the comfort theory of religious participation.

In **Figure 14.27** we see the results of a chi-square analysis. This analysis corresponds to the crosstabulation of frequency of prayer by gender that we showed in **Table 14.5**. (Note that regardless of whether you percentage by columns or rows, the crosstab's chi-square test will arrive at the same results.) In the first row, which is labeled "Pearson Chi-Square" (the formal name for the test), we see the results of the chi-square test in the columns to the right. The key inferential statistic to consider is the one listed under "Asymptotic Significance." This is the *probability value*, or ***p*-value**. It estimates the likelihood that we would see patterns as extreme as what we observed in our sample if there truly was no relationship between the two variables in our analysis (i.e., if the null hypothesis was true).

The chi-square test in this example has found that this chance is extremely low (<.001). This result tells us that the chance of seeing results this extreme if the null hypothesis was true is less than 1 in 1,000. That's way below 5 percent (.05). We therefore reject the null hypothesis that gender and frequency of prayer are unrelated in the U.S. adult population.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	119.111[a]	5	<.001
Likelihood Ratio	120.478	5	<.001
Linear-by-Linear Association	115.995	1	<.001
N of Valid Cases	2332		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 69.29.

Figure 14.27. Chi-Square Test for the Crosstabulation of Frequency of Prayer by Gender. This SPSS-generated table shows the results of a chi-square test accompanying a crosstab of the frequency of prayer variable (dependent variable) and gender variable (independent variable). As you can see on the right side of the first line, the *p*-value is less than .001, which is under the conventional threshold for statistical significance, .05. This means that there is a less than 1 in 1,000 chance that we would see these extreme differences between the prayer frequencies of men and women (as shown in the original crosstab in Table 14.5) if the null hypothesis was true. We can therefore reject the null hypothesis that no relationship exists between the two variables in the actual U.S. adult population.

When we sociologists feel reasonably certain that we should reject the null hypothesis, we say that the relationship

is **statistically significant**. But don't confuse this term with "important" or "meaningful." For one thing, the number of cases in the sample affects the results of null hypothesis tests. As your sample size grows larger, every statistic you generate from that sample will be statistically significant. Thus, even if the relationship between two variables is very weak, with a large number of cases you will obtain a statistically significant result for the chi-square test—or, for that matter, for any null hypothesis test. (See the sidebar **Understanding Null Hypothesis Testing** for more on the distinction between *statistical significance* and *practical significance*.)

Figure 14.28 presents a chi-square test to accompany the crosstab of frequency of prayer by educational attainment shown in **Table 14.5**. The *p*-value on the right tells us there is a .237 (23.7 percent) probability that we would see a pattern as extreme as what we observed in the earlier crosstab—which wasn't that extreme—if the null hypothesis was true. This figure is higher than .05, so we fail to reject the null hypothesis that the two variables are unrelated in the population. Our results are not statistically significant. In this case, we very well might have an unusual sample that's not accurately reflecting the situation in the larger population.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	6.787[a]	5	.237
Likelihood Ratio	6.724	5	.242
Linear-by-Linear Association	2.915	1	.088
N of Valid Cases	2334		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 46.56.

Figure 14.28. Chi-Square Test for the Crosstabulation of Frequency of Prayer by Educational Attainment. This chi-square test is for a crosstab of the frequency of prayer variable (dependent variable) and the highest year of school completed variable (independent variable). As you can see on the right side of the first line, the *p*-value is .237, which is over the conventional threshold for statistical significance, .05. This means that there is 23.7 percent chance that we would see these extreme differences between the prayer frequencies of people with different education levels (as shown in the crosstab in Table 14.5) if the null hypothesis was true. We therefore fail to reject the null hypothesis. There is a substantial chance that there is no relationship between the two variables in the actual U.S. adult population.

Understanding Null Hypothesis Testing

As we have noted at times throughout this textbook, researchers are generally interested in drawing conclusions not about their samples but rather about the populations their samples were drawn from. They

calculate frequencies, means, and other statistics from their samples to estimate the corresponding values in the population, what we call **parameters** (refer to the discussion in **Chapter 6: Sampling**). Unfortunately, sample statistics are not perfect estimates of their corresponding population parameters. This is because there is a certain amount of random variability in any statistic from sample to sample. For instance, in our GSS sample we found that 65.5 percent of people with a college degree pray several times a week or more, compared with 67.5 percent of those without a degree—a 2 percentage-point difference. But if we had taken multiple probability samples from the U.S. adult population, we would have likely gotten a slightly different frequency distribution for each one. The 65.5 percent result for college degree holders in the GSS sample might have been 67.1 percent in a second sample, and 64.3 percent in a third—even if these samples had all been selected randomly from the same population. This random variability in a statistic from sample to sample is called **sampling error**, or *random sampling error*. (Note that the term "error" here refers to random variability and does not imply that anyone has made a mistake—no one "commits a sampling error.")

Because of this inevitable variability that occurs across the probability samples we collect, whenever we observe a statistical relationship in our sample, it is not immediately clear that the same statistical relationship exists in the target population. As we noted, the GSS dataset showed a 2 percentage-point difference between the percentage of college-educated respondents who prayed several times a week or more and the percentage of non-college-educated respondents who did so. This small difference in prayer frequencies in the sample might indicate that there is a small difference between the corresponding frequencies in the population. But it could also be that no difference exists between the frequencies in the population—that the difference in the sample is just a matter of sampling error.

In fact, any statistical relationship in a sample can be interpreted in two ways:

- There is a relationship in the population, and the relationship in the sample reflects this.
- There is no relationship in the population, and the relationship in the sample reflects only sampling error.

The purpose of *null hypothesis testing* (also called *null hypothesis significance testing* or *statistical significance testing*) is simply to help researchers decide between these two interpretations. One interpretation is called the **null hypothesis** (often written as H_0 and read as "H-zero"). This is the idea that there is no relationship in the population and that the relationship in the sample reflects only sampling error. (A more informal way of stating the null hypothesis is that the sample relationship "occurred by chance.") The second interpretation is called the **alternative hypothesis** (often written as H_1). This is the idea that there *is* a relationship in the population, and that the relationship in the sample reflects this relationship in the population.

Again, *every* statistical relationship in a sample can be interpreted in either of these two ways. The relationship might have occurred by chance because of the sample we happened to collect, or it might reflect a real relationship in the target population. Researchers need a way to decide between these interpretations. Although there are many specific null hypothesis testing techniques (including the chi-square test mentioned earlier), they are all based on the same general logic and follow the same procedure:

- Assume for the moment that the null hypothesis is true: there really is no relationship between the variables in the target population.
- Determine how likely it is that the relationship we observed in the sample would occur if the null hypothesis were true.
- If the observed relationship would be extremely unlikely to occur by chance, then we *reject the null hypothesis* in favor of the alternative hypothesis. If it would not be extremely unlikely, then we *retain the null hypothesis*. (Note that social scientists also use this alternative phrasing for the latter idea: "We *fail to reject the null hypothesis*.)

Following this logic, we can begin to understand why our chi-square test in **Figure 14.28** concluded that there is no difference in the frequency of prayer of those with college degrees and those without. In essence, we asked the following question: "If there were really no difference in the U.S. adult population, how likely is it that we would find the small differences we saw in prayer frequency percentages across these two groups in our sample (including the 2 percentage-point difference in praying several times a week or more)?" Our answer to this question was that this sample relationship would be fairly likely if the null hypothesis were true. Therefore, we retained the null hypothesis—concluding that there is no evidence of a difference in frequency of prayer by educational levels in the population.

In turn, it makes sense why we rejected the null hypothesis in the chi-square test depicted in **Figure 14.27**, concluding that there is a relationship between frequency of prayer and gender in the U.S. adult population. Our null hypothesis test asked, "If the null hypothesis were true, how likely is it that we would find the large differences we saw in prayer frequency percentages across the male and female respondents in our sample (including the 18 percentage-point difference in praying several times a week or more)?" Our answer to this question was that this sample relationship would be fairly unlikely if the null hypothesis were true. Therefore, we rejected the null hypothesis in favor of the alternative hypothesis—concluding that there is a relationship between these two variables in the population.

Through null hypothesis testing, we determine the probability that we would detect a relationship this extreme in our sample if there really was no relationship in the target population. This probability is called the *p*-value. A low *p*-value means that a result this extreme in our sample result would be *unlikely* if the null hypothesis were true, and therefore we should reject the null hypothesis. A *p*-value that is not low means that a result this extreme in our sample would be *likely* if the null hypothesis were true, and therefore we should retain the null hypothesis.

The *p*-value is one of the most misunderstood concepts in quantitative research (Cohen 1994). Even professional researchers misinterpret it, and it is not unusual for such misinterpretations to appear in statistics textbooks! The most common misinterpretation is that the *p*-value is the probability that the null hypothesis is true—that the sample result occurred by chance. For example, a misguided researcher might say that because the *p*-value is .02, there is only a 2 percent chance that the result is due to chance and a 98 percent chance that it reflects a real relationship in the population. *This is incorrect.* To reiterate, the *p*-value is the probability we would observe a result as extreme as what we observed in our sample if the null hypothesis were true. So a *p*-value of .02 means that if the null hypothesis were true, a sample result this extreme would occur only 2 percent of the time.

How low must the *p*-value be before the sample result we observed is considered unlikely enough that we can reject the null hypothesis? In null hypothesis testing, this threshold is called α (**alpha**) and is conventionally set at .05. We reject the null hypothesis when our analysis tells us that the chance of observing a sample result this extreme if the null hypothesis were true would be less than 5 percent. If we meet this threshold, the result we obtained from our sample is said to be *statistically significant*. With a substantial degree of confidence, we can then generalize a statistically significant result derived from our sample to the target population.

We retain the null hypothesis when our analysis tells us that the chance of observing a sample result this extreme if the null hypothesis was true is 5 percent or greater. Importantly, this *does not* mean that the researcher accepts the null hypothesis as true. It only means that currently there is not enough evidence to reject it. This is the reason that researchers use language like "fail to reject the null hypothesis" and "retain the null hypothesis"—but never say they "accept the null hypothesis."

Why isn't there enough evidence to reject the null hypothesis? Recall that null hypothesis testing involves answering the question, "If the null hypothesis were true, what is the probability of a sample result as extreme as this one?" The answer to this question depends on two considerations: (1) the strength of the relationship in the sample, and (2) the size of the sample. Specifically, the stronger the observed relationship and the larger the sample, the less likely it would be for us to encounter this relationship in the sample if the null hypothesis was true. (Likewise, as the strength of the observed relationship and the size of the sample increase, the *p*-value—which is our measure of that likelihood—falls.)



Sadly, quantitative researchers who have not been blessed with the rigorous training you have received in this textbook come to the false impression that the *p*-value is the probability that the null hypothesis is true. Instead, it is the probability of obtaining the sample result if the null hypothesis really was true. Go out and wow your less fortunate friends with this knowledge. *Randall Munroe, via* XKCD

Our previous examples already gave us a sense of how the strength of the observed relationship directly affects the *p*-value and thus our decision to reject or retain the null hypothesis. The much stronger relationship between gender and frequency of prayer was highly statistically significant ($p < .001$, according to our chi-square test), whereas the weak relationship between educational attainment and frequency of prayer was well above alpha (the significance threshold we were using of .05). But the way that sample size also affects the *p*-value can complicate our analysis. Specifically, sometimes the relationship we observe in the sample can be very weak, but the result will still be statistically significant because our sample happened to be very large. Other times, the observed relationship can be fairly strong, but because our sample happened to be small, the result will not be statistically significant. This latter case is why we say we "fail to reject the null hypothesis," rather than outright "accepting" it. If we had gathered a larger sample, we very well may have concluded that the fairly strong relationship we observed was statistically significant. (Indeed, given that increasing our sample size always lowers our *p*-value in this way, we might say that the only time we can "accept" the null hypothesis is when our sample is the entire population!)

Table 14.7 gives you a rough sense of how relationship strength and sample size combine to determine whether

a sample result is statistically significant. The columns of the table represent three levels of relationship strength: weak, medium, and strong. The rows feature four sample sizes that can be considered small, medium, large, and extra large in the context of social scientific research. For each combination of relationship strength and sample size, the cells in the table tell us if null hypothesis testing would find the observed relationship to be statistically significant.

Table 14.7. The Statistical Significance of a Bivariate Relationship as Predicted by Relationship Strength and Sample Size

	Strength of the Bivariate Relationships		
Sample Size	**Weak**	**Medium**	**Strong**
Small (n = 20)	No	No	Maybe
Medium (n = 50)	No	Yes	Yes
Large (n = 100)	Maybe	Yes	Yes
Extra large (n = 500)	Yes	Yes	Yes

These predictions are only rough rules of thumb, but the breakdown in the table does show us clearly that (1) weak relationships based on medium or small samples are never statistically significant, and that (2) strong relationships based on medium or larger samples are always statistically significant. If you keep this logic in mind, you will often know whether a result is statistically significant based on the descriptive statistics alone. Developing this kind of intuitive judgment is extremely useful. It allows you to develop expectations about how your formal null hypothesis tests are going to come out, which in turn allows you to detect problems in your analyses. For example, if your sample relationship is strong and you're working with a medium-sized sample, then you would expect to reject the null hypothesis. If, for some reason, your analysis fails to reject the null, then you should double-check your computations and interpretations.

Table 14.7 illustrates another extremely important point that we hinted at earlier: a statistically significant result is not necessarily a strong one. Even a very weak result can be statistically significant if it is based on a large enough sample. For example, social scientists have found *statistically significant* differences between women and men in mathematical problem-solving and leadership ability, but as Janet Shibley Hyde (2007) points out, the differences speak to quite *weak relationships*—arguably, trivial ones. The word "significant" in this context can cause people to interpret these differences as strong and important—perhaps even important enough to influence the college courses they take. But that would be a gross exaggeration of what these weak but *statistically* significant relationships actually mean.

This is why it is important to distinguish between the *statistical significance* of a result and the *practical significance* of that result. **Practical significance** (also known as *substantive significance* or—in a treatment context—*clinical significance*) refers to the importance or usefulness of the result in some real-world context. Many relationships that social scientists uncover between variables are statistically significant—and may even be interesting for purely scientific reasons—but they are not practically significant. For example, a new intervention to reduce gang violence might show that the program produces a statistically significant effect, reducing the local rate of violent crime by 5 percent. Yet this apparent impact on violent crime rates might not be strong enough to justify the time, effort, and other costs of putting the intervention strategy into

practice—especially if easier and cheaper programs that work almost as well already exist. Although statistically significant, this result could be said to lack practical significance. As you might have guessed, determining substantive significance is inherently a subjective judgment, but it is vitally important in making sense of the true value and relevance of your research.

Understanding Crosstabs

Journals do not often include bivariate tables in the limited space devoted to articles, instead preferring that researchers describe these results within the text and focus any of their tables on presenting more complex multivariate analyses. Bivariate analyses like crosstabulations, however, can give us a richer sense of the relationships in our data, and you will occasionally see tables devoted to them in papers.

Figure 14.29 shows a published table including a series of bivariate crosstabulations and their associated chi-square tests. Eunice Kimunai, Salome Kapella-Mshigeni, and Peter Anderson (2016) examined surveys from 167 HIV-positive postpartum women of childbearing age, using data from the 2008–2009 Kenya Demographic and Health Surveys. Their analysis found that women with higher levels of education, literacy, or wealth were more likely to pursue certain types of care to prevent mother-to-child transmission of HIV.

	Unable to read at all		Able to read only part of sentence		Able to read whole sentence			
	n	%	n	%	n	%	χ^2	p
Drug Treatment for AIDS							5.34	.069
No	9	40.9	7	19.4	20	18.9		
Yes	13	59.1	29	80.6	86	81.1		
Got Results of AIDS During Antenatal Visit							7.15	.028
No	12	54.5	16	44.4	30	28.3		
Yes	10	45.5	20	55.6	76	71.7		
Tested for AIDS During Antenatal Visit							3.88	.144
No	9	40.9	15	41.7	28	26.4		
Yes	13	59.1	21	58.3	78	73.6		
Talk During Antenatal Visit Regarding AIDS Transmission From Mother to Child							1.89	.389
No	12	54.5	15	41.7	41	38.7		
Yes	10	45.5	21	58.3	65	61.3		
Counseling About Breastfeeding							1.81	.404
No	10	45.5	17	47.2	38	35.8		
Yes	12	54.5	19	52.8	68	64.2		
Things to Do to Prevent AIDS							2.143	.342
No	12	54.5	15	41.7	40	37.7		
Yes	10	45.5	21	58.3	66	62.3		

Figure 14.29. Table in Kimunai et al. (2016) Presenting Crosstabulations of Healthcare Behaviors by Level of Literacy. This table includes multiple crosstabulation analyses. Its columns cover the response categories for one independent variable (level of literacy), and its rows encompass multiple dependent variables relating to the types of health care that respondents pursued. The table also includes the chi-square statistics and *p*-values that assess the generalizability of the relationships detected between the independent variable and these dependent variables—with a separate relationship assessed in each row. *International Childbirth Education Association*

With a crosstabulation, one of your first tasks is to identify which variables are independent variables and which are dependent. The top of the table appears to list different degrees of literacy. There are three levels of literacy and two columns under each: "n" (count) and "%" (percentage). If we read the article, we learn that literacy is the independent variable in this analysis. The researchers wanted to see whether different levels of literacy affected the extent to which women sought out different types of health care. These healthcare options, the study's dependent variables, are listed in the rows of the table.

Even if we don't read the article text associated with a crosstab, however, we can usually figure out from the table itself which variables the authors are treating as the independent and dependent variables. First, what sorts of relationships between the variables listed in the table make sense theoretically? In our example, are we theorizing that AIDS drug treatment might cause variation in the respondent's level of literacy? Or that level of literacy might cause variation in the respondent's access to drug treatment? Just from a logical perspective, the latter is more likely.

Second, the way that the authors have presented their statistics in the table enable us to infer which variables they are treating as the independent and dependent variables. A crosstabulation table will usually be percentaged by columns if the independent variable appears in the table's columns, or by rows if the independent variable appears in the rows. How do we determine whether the table is percentaged by rows or columns? In the example crosstabs we looked at earlier, SPSS helpfully included "100%" subtotals for each category (in addition to the corresponding count subtotals), which told us exactly how the table was being percentaged. For space reasons, published tables often don't include percentage subtotals. In situations like these, we should add up all the percentages that are listed in one row of the table, and then add up all the percentages listed in one of its columns. If the percentages add up to 100 percent (accounting for any rounding) for the row you've examined, then the table is percentaged by rows; if they do so for the column, it is percentaged by columns.

In the example table, the first column adds up to 100 percent (40.9 + 59.1 = 100.0), so the table is percentaged by columns, and therefore the independent variable—level of literacy—appears in the columns. Knowing this, how would you interpret the "40.9" figure in the first row—what is 40.9 percent a percentage of? Remember that when you percentage by column, the denominator for the percentages (the percentage base) is the total for that column—in this case, the 22 women who indicated on the survey that they were "unable to read at all." (Although the table does not list a count subtotal, we know it is 22 because 9 women did not receive AIDS drug treatment—as listed in the first row—whereas 13 women did—as listed in the second row.) Thus, we would interpret the figure this way: 40.9 percent of the (22) women who are unable to read at all did not receive AIDS drug treatment (9 divided by 22 equals 40.9 percent).[7]

In a crosstab percentaged by columns, we should be comparing the percentages across the columns of a single row. Indeed, we see interesting patterns in the example table when we looks across its columns: in the second row, for instance, we find that a lower percentage of women who are unable to read at all (59.1 percent) are receiving drug treatment (the first column), compared with 80.6 percent of those who can read a part of a sentence (the second column) and 81.1 percent of those who can read whole sentences (the third column).

The final important thing to note is the column with our chi-square results, designated by the Greek letter *chi* ($\chi 2$) in the table. (See the earlier discussion of chi-square for details.) The chi-square test for each row tests the null hypothesis that there is no relationship between level of literacy and whatever dependent variable is listed on that row.

Let's look at the chi-square results on the first row of the table, which relate to whether the respondent received AIDS

7. Yes, we realize we boasted at the beginning of the chapter that we would not include formulas. Note, however, that we wrote out this formula in prose, without the use of mathematical symbols, which technically means it is not a formula (or at least that's our story).

drug treatment. At the end of this row, the table reports two inferential statistics: the value of the chi-square statistic (5.34) and its p-value (.069). (The chi-square statistic is used to determine the p-value; you can read more about how exactly it is calculated in a statistics textbook.) This chi-square statistic and this p-value are specifically testing the relationship between level of literacy (the table's one independent variable) and receipt of AIDS drug treatment (the dependent variable listed on the same row). The p-value of .069 tells us that if the null hypothesis were true (i.e., no relationship existed), there would be a 6.9 percent chance of seeing differences this extreme in how likely more and less literate women groups were to receive AIDS drug treatment. That's higher than the conventional statistical significance threshold of 5 percent, so we cannot reject the null hypothesis.

Going down the rows, we see that the p-value for being tested for AIDS during an "antenatal" (pre-birth) visit is also higher than .05. Therefore, we cannot reject the null hypothesis for the relationship between literacy level and that dependent variable, either. We can, however, reject the null hypothesis for the relationship between literacy level and receiving the results of an AIDS test at an antenatal visit. These two variables might actually be related in the larger population.

It's worth a reminder at this point that null hypothesis testing should be conducted only on probability samples. The survey for the example study was conducted using probability-sampling methods, but you will frequently see null hypothesis testing done on nonprobability samples. Avoid that temptation in your own research. If you don't have faith that your sample is a random sample of your population of interest, any statistics you generate can't be generalized—even if your p-values are reassuringly low.

How do we summarize the results of a crosstab in prose? Null hypothesis testing complicates how we describe bivariate relationships, so we'll cover this topic in the separate sidebar **Interpreting Inferential Statistics**. Note that many of the suggestions in this sidebar can be applied to other bivariate analyses besides crosstabs.

Interpreting Inferential Statistics

When reporting the results of a crosstab, the first thing to keep in mind is whether or not the null hypothesis test you used found a statistically significant relationship. If there is no statistical significance, you do not want to emphasize the patterns you observed in the sample, because you do not have confidence that those patterns actually exist in the target population (i.e., the analysis failed to reject the null hypothesis). You should state upfront that the relationship between the independent variable and dependent variable is not statistically significant, including a parenthetical notation to specify the threshold you used for statistical significance—for example, ($p > .05$). Another way of conveying the same thing (i.e., nonsignificance) is to simply state that the dependent variable does not differ across categories of the independent variable. You might note a few key statistics from your crosstab to describe your results, but it is not worthwhile to go into much depth because any patterns you point out may not actually apply in the target population. If you do mention any observed differences, you will also want to add the caveat that any apparent relationship is likely to be due to random sampling error.

If your analysis found the relationship to be statistically significant (i.e., the null hypothesis was rejected), you can go ahead and interpret the results using the step-by-step GEE approach (Miller 2015). In the generalization step, you can explicitly state that the relationship is significant, or you can add the word "significantly" when describing any patterns—for example, "We find that older respondents pray significantly more often than younger respondents." Include the p-value (or the significance threshold that the p-value reached) in parentheses at the end of this sentence—for example, ($p < .001$) or ($p = .027$). Since the relationship is

statistically significant, you should mention notable statistics that illustrate the pattern you found (as part of the "example" step). Provide any measures of the strength of the observed relationship, such as percentage-point differences across key cells in a crosstab. However, do not include every statistic listed in the table. Finally, in the "exceptions" step, mention any statistics that violate the pattern you mentioned earlier—for instance, if middle-aged respondents do not fall into the pattern that is seen across the oldest and youngest groups of respondents in your sample.

A special case is when your p-value is slightly above 0.05. For these "marginally significant" cases ($p < .1$), you might want to mention the sample size and point out that the observed relationship might have been statistically significant if the sample had been larger (see the discussion of sample size in the sidebar **Understanding Null Hypothesis Testing**).

When you are writing for nonacademic audiences, you can drop the references to p-values, as lay readers will not understand them. Instead, try to break down the concepts into more accessible terms. For statistically significant results, you can state that you have evidence to believe that the relationship between the two variables is not due to chance. You can also specify that the chances of observing an association this strong in your sample if there were no real relationship between the two variables is "less than 1 in 20" (for a p-value right under the conventional .05 threshold) or "less than 1 in 1,000" (for a p-value under .001).

For nonsignificant results, you should never talk about the existence of any patterns in your data: you do not have strong evidence that these patterns are not due to sampling error. For lay audiences, state that the dependent variable does not differ across the categories of the independent variable, and point out that any patterns or differences observed could have easily occurred through chance alone.

Finally, remember that in any discussion of bivariate results, you should not state that the relationship is causal—for instance, that a change in the independent variable "causes" a change in the dependent variable, or that you have observed the "effect" of the independent variable on the dependent variable. Instead, use more neutral words like "associated," "correlated," "related," and "linked" to describe the connections between variables. Avoid talking about "consequences" or "effects" and describing the independent variable as "causing" or "affecting" the dependent variable—these phrasings imply causality. More broadly, describe your results in a way that acknowledges that alternative factors may be at play, since your bivariate analysis cannot account for other possible explanations for the patterns between the independent and dependent variables that you see.

Key Takeaways

1. For a straightforward comparison that is easy to interpret, place your independent variable in the columns of your crosstab and your dependent variable in the rows, and percentage by columns.
2. Analyze the strength, pattern, and direction (if using ordinal-level variables) of the relationship depicted by your crosstab.
3. To generalize the results of your crosstab to the population of interest, you first need to have collected a probability sample. A strong relationship and a larger sample size makes it more likely that your results will be statistically significant, meaning that you can reject the null hypothesis that any relationship you observed was merely due to a fluke sample, rather than an actual relationship in the target population.

4. You can use the chi-square test to test the statistical significance of the patterns of frequencies you observe in a crosstab. If the *p*-value the test generates is lower than your alpha (the threshold for statistical significance, typically .05), then you have confidence that the observed results are not occurring due to sampling error—a sample that, by chance, does not reflect the relevant characteristics of the target population.

Exercises

1. Write a few sentences to describe the bivariate relationship between frequency of prayer and gender that is depicted in the crosstab in **Figure 14.23**. Then compare your write-up to the suggested answers.
2. Write a few sentences to describe the eight bivariate relationships depicted in the crosstabs in **Table 14.5** and **Table 14.6**. Then compare your write-ups to the suggested answers.

Suggested Answers

- Our hypothesis for the crosstab in **Figure 14.23** was that women would have a higher frequency of prayer than men would. Our analysis of the 2018 GSS sample of U.S. adults supported this hypothesis, finding that 36 percent of women prayed several times a day, 15 percentage points higher than the rate for men. Conversely, 23 percent of men said they never prayed, 13 percentage points higher than the rate for women.
- For **Table 14.5** and **Table 14.6**, we hypothesized that the higher a respondent's age, the higher their frequency of prayer and the higher their frequency of attendance at religious services. In the GSS sample of U.S. adults, older respondents prayed and attended religious services more often, in line with this hypothesis. For example, 77 percent of respondents 65 and older prayed several times a week or more often, compared with 53 percent of those ages 18 to 39, a difference of 24 percentage points. Similarly, 46 percent of those 65 and older attended religious services more than three times a month, compared with 28 percent of those ages 18 to 39, a difference of 18 percentage points.
- We hypothesized that women would have higher frequencies of prayer and attending religious services than men would. In line with the first hypothesis, women prayed significantly more often than men ($p < .001$): 75 percent of women prayed several times a week or more often, compared with 57 percent of men, an 18 percentage-point difference. (We have included null hypothesis testing results in this particular write-up because they were provided earlier in **Figure 14.27**.) In support of the second hypothesis, 39 percent of women attended religious services more than three times a month, compared with 31 percent of men, an 8 percentage-point difference.
- Our hypotheses were that nonwhites would have a higher frequency of prayer and a higher frequency of attending religious services than whites would. We observed these patterns in the GSS sample, with 73 percent of nonwhites saying they prayed several times a week or more often, a rate 9 percentage points higher than for whites, and 41 percent saying they attended religious services more than three times a month, a rate 8 percent points higher than for whites.
- Our predictions at the outset of this analysis were that the lower the respondent's educational level, the

higher their frequency of prayer and the higher their frequency of attending religious services. Of respondents with and without college degrees, 66 percent and 68 percent, respectively, prayed several times a week or more often. This 2 percentage-point difference across the two education groups was not statistically significant ($p > .05$); frequency of prayer does not differ by educational attainment. (These hypothesis testing results correspond to the details provided in **Figure 14.28**.) Our other hypothesis was also not confirmed: college graduates were actually more likely to attend religious services, with 40 percent attending more than three times a month, which was 7 percentage points higher than the rate for respondents without college degrees.

14.4. References

Babbie, Earl. 2016. *The Practice of Social Research*. 14th ed. Boston: Cengage Learning.

Cohen, Jacob. 1994. "The Earth Is Round (p < .05)." *American Psychologist* 49(12):997–1003. doi: 10.1037/0003-066X.49.12.997.

Glock, Charles Y., Benjamin B. Ringer, and Earl R. Babbie. 1967. *To Comfort and to Challenge*. Berkeley: University of California Press.

Harding, David J., Jeffrey D. Morenoff, and Jessica J. B. Wyse. 2019. *On the Outside: Prisoner Reentry and Reintegration*. Chicago: University of Chicago Press.

Hyde, Janet Shibley. 2007. "New Directions in the Study of Gender Similarities and Differences." *Current Directions in Psychological Science* 16(5):259–63. doi: 10.1111/j.1467-8721.2007.00516.x.

Kimunai, Eunice, Salome Kapella-Mshigeni, and Peter Anderson. 2016. "Mother-to-Child HIV Transmission, Literacy, Ethnicity, Education, and Wealth in Kenya." *International Journal of Childbirth Education* 31(2):8–17.

Lee, Bob, Yiwei Chen, and Lynne Hewitt. 2011. "Age Differences in Constraints Encountered by Seniors in Their Use of Computers and the Internet." *Computers in Human Behavior* 27(3):1231–37. doi: 10.1016/j.chb.2011.01.003.

Miller, Jane E. 2015. *The Chicago Guide to Writing about Numbers*. 2nd ed. Chicago: University of Chicago Press.

Murphy, Tom. 2020. "U.S. Survey Finds Smaller Decline in Medical Bill Worries." *San Francisco Chronicle*, February 12.

Nigam, Minali, and Elizabeth Cohen. 2020. "French Study Finds Hydroxychloroquine Doesn't Help Patients with Coronavirus." CNN, April 15.

Schor, Elana. 2020. "White Supremacist Propaganda Spreading, Anti-Bias Group Says." *Associated Press*, February 11.

15. MATERIALS-BASED METHODS



Are female and male athletes treated equally? At least at the college level, you might think that the answer would be an easy "yes." After all, it's been many decades since the passing of Title IX, a 1972 civil rights law that prohibits sex discrimination in education, including athletics. That legislation has been credited for bringing about substantial increases in female participation on high school and college sports teams. More broadly, the culture has changed. In widely televised sports like basketball, golf, and tennis, women athletes have gained greater visibility and sometimes even celebrity status.

Sociologists have used their skills in data collection and analysis to address this question of gender discrimination in sports. And they've found creative ways to measure such discrimination. For instance, Jo Ann Buysse and Melissa Sheridan Embser-Herbert (2004) analyzed media guides published by college athletics departments. Their examination of existing texts—what we call a **content analysis**—showed how colleges maintain traditional definitions of femininity by visually presenting women athletes not as strong and athletic, but passive and overtly feminine. Another study examined gender discrimination by reviewing televised coverage of men's and women's sports (Messner, Duncan, and Jensen 1993). The content of this content analysis was the running commentary by announcers—that is, treating their verbal exposition as the "text" to be studied. Although the two studies we have just described analyzed very different media, what they have in common is that they both draw on existing materials. As it happened, the content analysis of sports coverage also found gender bias: broadcasters tended to infantilize women athletes and express ambivalence about their accomplishments.



Content analyses of how women athletes are portrayed across various kinds of media have provided evidence that gender discrimination persists in sports. *Gaby Lopez, via Pexels*

In this chapter, we explore **materials-based methods**, also referred to as *unobtrusive research*. Rather than conducting their own interviews, surveys, or observations, materials-based researchers draw upon materials that already exist. Furthermore, they normally need to interact only with those materials, not directly with the people who created them. It may seem strange that sociology, a discipline dedicated to understanding human social behavior, would employ a methodology that requires no interaction with human beings. But people create plenty of evidence of their behaviors. They write books, produce movies and TV shows, consume goods, walk on sidewalks, and lie on the grass in public parks. All these activities leave something behind—printed papers, recorded shows, trash, and worn paths. These are all potential sources of data for a materials-based project.

In presenting materials-based research as a distinct methodological category, Deborah Carr and her collaborators (2020) identify the many kinds of materials that researchers can draw upon, which we elaborate on in the following list:

- **Expert analyses**, which are often in the form of books and articles written by academics, scientists, and pundits.
- **Reports** written by governments or other organizations to summarize relevant information on a particular topic.
- **News articles and broadcasts** produced by media organizations about current and historical events and issues of interest.
- **Documents and records** collected by governments and other organizations, such as census forms, birth

certificates, marriage records, death certificates, treaties, police records, laws, and emails and letters sent by organizational staff. (Note that many of these government documents and records can be obtained by researchers through Freedom of Information Act requests at the U.S. federal level and similar legislation in other contexts.)
- **Artifacts of popular culture**, such as novels, songs, television shows, movies, and comic books. (Note that sometimes researchers refer to the materials analyzed in materials-based research as **cultural artifacts**.)
- **First-person accounts of events** and other individual commentaries, such as diary and journal entries, correspondence (from snail-mail letters to emails to direct messages), memoirs, oral histories, and blog and social media posts. (We'll have more to say about digital content later in this chapter.)
- **Physical materials** from toys and political campaign materials, to bumper stickers and yard ornaments, to foodstuffs and refuse—anything that could be considered a physical cultural artifact that says something about the individuals or communities that created it.

We mentioned how government records like census forms and birth certificates could be utilized in content analyses. It's worth noting that these kinds of materials are often the basis for secondary data analysis, which we discussed in **Chapter 14: Quantitative Data Analysis**. The line between secondary data analysis and materials-based research can get blurred, and it is often more a matter of emphasis. For instance, a secondary data analysis would aggregate the responses on forms and certificates to produce statistics—an age breakdown for the U.S. population based on census records, so-called vital statistics on birth, marriage, and death rates, and so on. By contrast, a content analysis might focus on the wording of the forms and records themselves—perhaps using techniques for analyzing qualitative data.

In this chapter, we focus on the widely used approach to materials-based research known as content analysis. You should note, however, that sociologists also use historical-comparative methods to analyze materials in a more sweeping way.

Opening chapter image credit: *Ron Lach, via Pexels. Adapted by Bizhan Khodabandeh.*

15.1. Content Analysis

Learning Objectives

1. Define content analysis.
2. Describe qualitative and quantitative strategies employed in content analysis.
3. Understand how to present the results from content analysis.

As we noted earlier, *content analysis* is a materials-based research method that focuses on texts and their meanings. Sociologists use a more expansive definition of "text" than the word typically has. In a research context, the content being analyzed is essentially any recorded communication. This would obviously include actual written copy, such as news articles or email messages, but we can consider content that we might see or hear—such as a speech, dance performance, television show, advertisement, or movie—to be "texts" as well. **Table 15.1** provides some examples of the kinds of data that sociologists have studied using content analysis techniques.

Table 15.1. Examples of Content Analysis

Data	Research Question	Authors and Year of Publication
Spam emails	Is the frequency of apologies from advertisers in unsolicited emails different in the United States and Korea?	Park, Lee, and Song 2005
James Bond films	How are women portrayed in James Bond films, and what broader lessons can be drawn from these portrayals?	Neuendorf et al. 2010
Console video games	How is male and female sexuality portrayed in best-selling console video games?	Downs and Smith 2010
News articles	What kinds of framing contests emerged in the news about Colin Kaepernick's NFL protests against racial oppression and police brutality?	Boykoff and Carrington 2020
Pro–eating disorder websites	What characteristics do websites that endorse and support eating disorders share, and what messages do they communicate to users?	Borzekowski et al. 2010

One thing you might notice is that the data sources described in this table are **primary sources**. As you may remember from **Chapter 5: Research Design**, primary sources are original works representing first-hand experiences, often written by individuals who were present at a noteworthy event or had relevant experiences. Primary sources that

could be studied through content analysis include personal journals, emails, letters, government documents, speeches, television commercials, social media posts, and news articles published at the time of an event of interest.

Although content analysis usually focuses on primary sources, there are also examples of studies that use **secondary sources** (which draw upon primary sources for information, such as academic publications, biographies, and news articles that review other media or research) and **tertiary sources** (which summarize the results of secondary sources). With these sources, researchers might be interested in how the person or persons who generated the secondary or tertiary source reached their conclusions about the topic in question, or how their decisions about presenting information might shape people's understandings. For example, Myra Marx Ferree and Elaine Hall (1990) conducted a content analysis of introductory sociology textbooks to learn how students were being taught sociology. As part of their study, the researchers examined the images being presented to students and what messages they conveyed. They concluded that "women were not represented in numbers proportionate to their distribution in the population" and people of color "would have been numerically underrepresented" in images had the textbooks not included chapters specifically on race (Ferree and Hall 1990:529, 528).

Sometimes students new to research methods struggle to grasp the difference between a *content analysis* of primary scholarly literature and a *review of literature* on a sociological topic. As we discussed in **Chapter 5: Research Design**, a literature review examines sources to understand what we know and don't know about a particular topic. These sources are typically peer-reviewed, written by trained scholars, and published by an academic journal or press. They are primarily papers that elaborate on scholarly theories or present empirical research conducted using accepted techniques of data collection and analysis for the discipline. The researcher synthesizes these sources to arrive at some conclusion about social scientists' overall knowledge about a topic—often as a prelude to designing their own study on that topic.

A content analysis of research studies would use its sources to pursue very different questions. In short, it would be a "study of the studies" as opposed to a "review of studies." For example, a content analysis of scholarly literature might ask whether the top-ranking journals in a particular discipline disproportionately publish work by men. The researcher might gather a sample of articles published in different journals and count their authors by gender (although this may be a tricky prospect if relying only on names to indicate gender). Another content analysis of academic studies might examine whether and how the topics that scholars study go in and out of style. A researcher tackling this research question could look at the articles published in various journals, code the topics they cover, and see if there is any change in the most popular topics over the years. Unlike with literature reviews, the researchers in these examples are not examining the content of their sources to identify a gap in the literature and design a study around it. Instead, they are looking to learn about what the publication of these articles says about the scientific community that published them.

Qualitative versus Quantitative Approaches to Content Analysis



Kimberly Neuendorf and her colleagues (2010) conducted a quantitative content analysis of 20 James Bond films to assess the portrayals of 195 female characters. *MachoCarioca, via Wikimedia Commons*

Content analysis can be quantitative or qualitative, and often researchers will use both strategies to strengthen their investigations. **Quantitative content analysis** focuses on variables whose characteristics can be counted. For example, Kimberly Neuendorf and her colleagues (2010) reviewed 20 films featuring the fictional British spy James Bond. They examined the portrayals of women across the films—195 female characters in total—and tracked particular details about each character. For example, a variable they called "role prominence" assessed whether the female's part was minor, medium, or major. Other variables counted the number of times weapons were used by and against a female character. The approach in such a quantitative content analysis uses the same techniques we covered in **Chapter 14: Quantitative**

Data Analysis. Neuendorf and her coauthors, for instance, calculated frequencies for their three categories of "role prominence," finding that 52 percent of female roles were minor, 30 percent medium, and 17 percent major. They generated both univariate and multivariate statistics—for instance, determining that 25 percent of female characters had a weapon used against them, and their own use of weapons was related to their levels of sexual activity within the films.

In **qualitative content analysis**, the aim is to identify themes in the text and examine the underlying meanings of those themes. Tony Chambers and Ching-Hsiao Chiang (2012) used such an approach in their content analysis of open-ended comments in the National Survey of Student Engagement. First, the researchers identified passages in the comments where students appeared to be raising issues standing in the way of their academic success. Through the coding process (see Chapter 11: Qualitative Data Analysis), they settled on certain key themes across these passages, which could be broadly grouped into categories of academic needs, the campus environment, financial issues, and student services. Their approach was inductive, in that the researchers did not determine these themes beforehand, but rather allowed them to emerge during the coding process. Ultimately, this analysis allowed Chambers and Chiang to highlight certain disconnects between the goals of colleges and universities and students' actual experiences.

Note that both quantitative and qualitative approaches can easily be used in the same content analysis. For example, Vaughn Crichlow and Christopher Fulcher (2017) analyzed coverage in the *New York Times* and *USA Today* about the deaths of three African American men while they were in police custody: Eric Garner, Michael Brown, and Freddie Gray. The content analysis focused on the quotes of experts (public officials and pundits) that appeared in these news articles. First, the researchers examined the quotes to generate themes, and then they tallied the numbers of quotes that fit the noteworthy themes they identified (as shown in this example, content analysis can move rather seamlessly from inductive qualitative analysis to deductive quantitative analysis). As shown in Figure 15.1, the analysis found that the experts quoted across these articles rarely discussed strategies to reduce police shootings or to improve police-community relations in communities of color.

Table 3 Experts' quotes on reducing police shootings and improving police–community ($n = 120$)

	Incident 1 (Staten Island)		Incident 2 (Ferguson)		Incident 3 (Baltimore)	
	n	%	*n*	%	*n*	%
Reduce shootings	0	0	0	0	0	0
Improve relations	9	23	8	20	9	23
Both	1	2	1	3	1	2
Not mentioned	30	75	34	77	30	75

Figure 15.1. Expert Quotes about Police Shootings in the New York Times and USA Today: A Content Analysis by Crichlow and Fulcher (2017). Vaughn Crichlow and Christopher Fulcher used both qualitative and quantitative approaches in their content analysis of news articles about the deaths of three African American men while in police custody: Eric Garner in Staten Island, New York; Michael Brown in Ferguson, Missouri; and Freddie Gray in Baltimore. The researchers first qualitatively analyzed the quotes of experts across these articles to inductively identify a set of common themes, and then they went back to the data to count the number of instances when key themes appeared. The analysis depicted in the table shows that the vast majority of expert quotes following each of these police-involved deaths did not mention efforts to reduce police shootings or improve police-community relations (Crichlow and Fulcher 2017:176).

One important thing to note about content analysis is the critical importance of developing a comprehensive and detailed codebook. For the study of James Bond films, for instance, the eight coders assigned values to the study's variables drawing on a codebook the authors had created well before any analysis began. Especially when many people are involved with coding, it's important to establish a shared understanding of what exactly the codes mean to avoid divergent interpretations of the same data and maintain *inter-coder reliability* (a topic we will return to).

Defining the Scope of a Content Analysis



Jules Boykoff and Ben Carrington (2020) used content analysis to examine media coverage of NFL player Colin Kaepernick's 2016 kneeling protests against racialized oppression and police brutality. *Lars Plougmann, via Flickr*

For many studies that rely on content analysis, setting boundaries on what content to study is challenging because there is just so much content available to researchers. Consider a study by Jules Boykoff and Ben Carrington (2020) that analyzed media coverage of American football player Colin Kaepernick's protests in 2016. By kneeling during the playing of the national anthem at numerous NFL games, Kaepernick sought to make a public statement about ongoing police brutality against African Americans. For their study of reactions to these protests, researchers examined "the media framing contests that emerged between Colin Kaepernick and his supporters on one side and his detractors on the other, including President Donald Trump" (Boykoff and Carrington 2020:832).

If the idea is to study how the media presents a particular issue relating to racial inequality, however, we need to be very specific about what we mean by "media." The term covers a lot of territory, even if we restrict it to just "mass media"—news articles, radio and television broadcasts, and the like. Conceivably, we could analyze reports about Kaepernick in hundreds of sources across a wide range of media platforms. Clearly, however, that is not a realistic option for any study, even a well-funded one.

To make our content analysis feasible, we need to precisely define the **scope conditions** of our study. In **Chapter 3: The Role of Theory in Research**, we talked about how all theories have scope conditions, which tell us where the theory can and cannot be applied. Among other things, a theory is constrained (*delimited*) based on the contexts where it has been studied empirically. For instance, a theory developed from data in the United States may not necessarily apply in other societies. When we are designing an empirical study, we need to think about the reverse consideration—what can we feasibly study? How should we set the boundaries of our data collection? These are the scope conditions of our study.

For their study, Boykoff and Carrington decided to study print and online stories from newspapers—thereby reducing the vast category of "media" to the smaller (and dwindling) subcategory of "newspapers." But which newspapers? In 2018, when Boykoff and Carrington were working on their analysis, there were 1,279 daily newspapers operating across the country. Even though this was fewer than in previous years, examining multiple articles in all of these publications would clearly have been an overwhelming task. Instead, the researchers decided to choose the four national newspapers with the highest circulation numbers—*USA Today*, the *Wall Street Journal*, the *New York Times*, and the *Los Angeles Times*—plus the seventh-ranked paper, the *Washington Post*. Their rationale for the first four had to do with the national reach and influence of these papers. They chose to include the lower-circulation *Washington Post* as well because of its in-depth coverage of issues and its numerous online articles. For any content analysis, you should be specific in a similar way about your rationale for including or excluding certain sources. Always provide a clear and compelling justification—detailed at length in your paper's methods section—for these methodological choices, rather than just saying these were the most convenient sources to analyze. Convenience can be an important consideration, but you should be able to speak thoughtfully about the benefits and drawbacks of the boundaries you set.

Even after Boykoff and Carrington settled on five newspaper sources to analyze, they still had another set of methodological decisions to make: what particular time period should they set for the Kaepernick-related articles they would analyze? Clearly there would be no coverage before the start of Kaepernick's protests on August 14, 2016, so they could easily set a starting point for their analysis. The ideal ending point was not self-evident, however, given that newspapers obviously don't coordinate with each other about when to stop coverage of a story. For their study, the researchers chose to examine news articles within a two-year period ending on August 14, 2018. By that time, they reasoned, the controversy had diminished, and no other NFL team had signed Kaepernick. This brings up another important point about setting scope conditions for a study: sometimes the decisions can be more or less arbitrary. While a two-year period is reasonable and justifiable, you could imagine how other researchers might make a decent case for a longer or shorter period of time. As a researcher, you will need to trust your own judgment and be able to defend your decisions, even if they are a matter of personal preference.

Now that Boykoff and Carrington had settled on a time period for their study, were they finally done? No. They still had to decide whether they should study all articles published during this period or be more selective. They decided to apply certain **inclusion criteria** to make it more likely that the articles they analyzed would contain material relevant to their research question of how the media framed Kaepernick's actions. In their first sweep, they searched the archives of each newspaper for the terms "Kaepernick" and "protest." Then they narrowed that list to just those articles that were at least five paragraphs long and that mentioned Kaepernick somewhere in the first five paragraphs of text, reasoning that these articles would be more likely to say something substantive about the controversy. By defining the scope of their content analysis in these ways, Boykoff and Carrington ensured that closely reviewing their materials was something they could do with a reasonable amount of time and effort. At the same time, they maximized the likelihood that their analysis would capture the vast majority of newspaper articles that put forward a useful (for their research purposes) perspective on the protests.

Let's look at another example. We previously mentioned a published content analysis of gender and racial representations within the images published in introductory sociology textbooks (Ferree and Hall 1990). To identify sources for their analysis, the authors consulted the 1988 volume of *Books in Print*, a comprehensive database of books

published in a particular year. They searched the database with the keywords "sociology" and "sociological" and obtained the most recent editions of textbooks whose titles contained those words.

Through this process of culling, the authors dramatically reduced the number of textbooks they had to analyze. However, they had another decision to make: should they analyze all the content within each textbook? For their research purposes, they decided to ignore the written content in the textbooks and instead focus on the accompanying images. Their reasoning—which, as with any methodological decision, could be debated—was that the photographs and illustrations vividly captured "the currently acceptable conceptualization of race and gender, as constructed in introductory sociology textbooks by authors and publishers" (Ferree and Hall 1990:501). Because of this specification, the researchers were able to eliminate textbooks that did not contain any photographs or illustrations. Their final sample included 33 introductory sociology textbooks with a total of 5,413 illustrations, all of which were subjected to analysis.

Even after you have defined the overall scope of your content analysis, you can be even more selective about the texts you examine by applying a sampling process. Say that your research question is, "How often do presidents of the United States mention domestic issues as opposed to international issues in their State of the Union addresses?" Luckily, a comprehensive list of State of the Union addresses that includes their full text is easy to locate on the internet. With all these sources readily available to you, you might decide you want to analyze every single speech. In this case, you would not need to select a sample. An examination of all of the possible elements in a population is called a **census**. The content analysis of James Bond films we mentioned earlier fits this category (Neuendorf et al. 2010). The authors conducted a census of every James Bond film available to them—from *Dr. No* (1962) starring Sean Connery, to *Die Another Day* (2002) starring Pierce Brosnan. (Their analysis went up to 2005, so they did not cover any of the Daniel Craig films.) Given the relatively small number of films they needed to analyze—just 20—the researchers were able to include all the elements in their target population.

Indeed, what resources you have available to you will often dictate whether you can study all possible sources or need to draw a smaller sample. Even though it is easy to identify and locate all 200-some State of the Union addresses, for instance, you may not have the time or energy to analyze every single one. For the hypothetical study mentioned earlier, let's say you look over some of the State of the Union addresses, and given how long and dense many of them are, you decide you can analyze only 20 of them. So, how do you decide which 20 to look at?

You can select a sample by using either probability or nonprobability sampling methods, as described in **Chapter 6: Sampling**. Let's walk through the process of drawing a simple random sample of 20 State of the Union addresses:

1. As of this writing, there have been 233 State of the Union addresses. So you should start by creating a list of all 233 addresses and then numbering them from 1 to 233.
2. Generate 20 random numbers in the range of 1 to 233 using a website such as random.org.
3. Go down your list of random numbers, choosing State of the Union addresses whose numbering in your original list of addresses corresponds to each random number. Ignore any duplicate numbers (you'll need to generate a new random number to replace it) and keep going until you have selected 20 addresses.

A key advantage of using **probability sampling** techniques, as you might remember, is that they allow you to generate statistics that estimate the population values—giving you a sense of how certain characteristics are distributed and whether certain relationships exist in the actual population. So if you selected your 20 State of the Union addresses through a simple random sampling procedure (or some other probability sampling technique), you would be able to generalize the results from your sample of 20 addresses to the population of all 200-some addresses. If, for example, you found that text relating to international issues accounted for a small percentage of the total text across your 20 State of the Union addresses, you could with some confidence say that the same percentage applied to discussions of international issues across all State of the Union addresses.

Content analyses can also rely on samples that were not drawn using probability sampling techniques. As we described

in **Chapter 6: Sampling**, the best choice among nonprobability sampling techniques is usually **purposive sampling**, which uses theory to guide the researcher in deciding what elements to include. For the example research question we've been considering that examines the extent to which presidents discuss international and domestic issues during their State of the Union addresses, a purposive sample might include the most recent 10 addresses delivered in wartime and the most recent 10 delivered in peacetime. This sampling choice would allow you to understand how the prevailing context of war or peace influences how presidents talk about international issues in their public statements. Note, however, that your nonprobability sample here would not allow you to make statistical estimates of population parameters for all 233 addresses.

Even if you choose to study all the units in your population of interest, you may still have a use for sampling techniques. Researchers often select smaller samples from their populations to develop coding schemes. For example, Dina Borzekowski and her colleagues (2010) did a comprehensive internet search to identify 180 active websites that endorsed and supported eating disorders. Before they started going through all 180 websites, however, the researchers reviewed the existing literature and used it to develop a preliminary coding scheme that described a number of relevant variables that they might measure within the data. They then tested that proposed scheme on a random sample of 25 of the 180 websites. The data from this **pilot testing** helped the researchers develop a codebook with detailed coding guidelines, which the team of six researchers then used to code the entire sample in a consistent fashion.

A similar approach of testing coding schemes with samples can be used for qualitative content analysis. After compiling their sample of newspaper articles, Boykoff and Carrington (2020:833) selected a random 30-article subsample to help them identify the frames that their media sources applied to coverage of Colin Kaepernick's kneeling protests. One of the study's authors and a research assistant carefully read all articles in that subsample and inductively generated a list of relevant frames. After consulting with one another, the researchers boiled down their list to seven predominant frames and three competing frames, which they then used to code the entire sample of 301 articles.

Coding in Quantitative Content Analysis

When researchers code a text for a content analysis, they are looking for two types of content: manifest and latent content. **Manifest content** has a more or less obvious significance to it. It presents a surface-level meaning that can be readily discerned. **Latent content** refers to the underlying meaning of the surface content we observe. Coding manifest content typically involves a highly structured process—think of the precise ways, for example, that the study of James Bond films we mentioned earlier tallied the cases when women were targets of or wielders of weapons (Neuendorf et al. 2010). Qualitative researchers, however, are usually not interested in manifest content in itself, but rather use it as a stepping stone to gain deeper understanding, which they pursue with the coding of latent content.

Say that we are studying advertisements for kitchen utensils that have been published in magazines over the past several decades. One type of manifest content to code across these ads might be the stated purpose of each featured utensil. For instance, the text in the ads might emphasize how the utensil—say, a potato masher—makes preparing mashed potatoes a cinch for Thanksgiving dinners. Or it might focus instead on the general versatility and efficiency of that potato masher in dealing with russet, red, and Yukon Gold potatoes, or even the dreaded yam. We could also make note of how users of the utensils are depicted within the ads, using codes like "entertaining at home" when we see someone in the ad using the utensil to cook for a large gathering (refer to **Chapter 11: Qualitative Data Analysis** for a review of how to write up codes). In turn, one set of latent codes that could emerge from these manifest codes would be our assessment of the lifestyles that the ad—by playing up certain features of the utensil—is promoting. For instance, we might see a shift over the years in the stated purpose of the utensils and the depictions of its users across the ads we are analyzing: from an emphasis on utensils designed to facilitate in-home entertaining to those designed to maximize efficiency and minimize time spent in the kitchen. We might theorize that this shift reflects a corresponding shift in how (and how much) people spend time in their homes. (See **Video 15.1** for one take on this woefully understudied topic of kitchen utensils.)

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://viva.pressbooks.pub/sociology-research-methods/?p=526#oembed-1

Video 15.1. What Kitchen Utensils Say about Society. In the first part of this segment from the PBS documentary *People Like Us*, satirist Joe Queenan riffs on how kitchen utensils can serve as markers of social class—in essence, an argument that one might explore through the latent content of kitchen utensil advertisements, as we describe.

To record the observations we make during content analysis, we typically rely on a **code sheet**, sometimes referred to as a *tally sheet*. Code sheets allow us to apply a systematic approach to our analysis of the data. For instance, let's say we want to conduct a content analysis of kitchen utensils—this time, not the advertisements about them, but the utensils themselves (remember that any "text"—even physical objects—can be studied in a content analysis, so long as the object conveys meanings we can analyze.) We happen to have access to sales records for kitchen utensils over the past 50 years. Based on these records, we generate a list of 50 utensils, the top-selling utensils in each year. For each utensil, we use our code sheet (as shown in **Table 15.2**) to record its name, culinary purpose, and price in current dollar amounts (note that adjusting for inflation is crucial whenever you compare monetary amounts across years). We might also want to make some qualitative assessments about each utensil and its purpose—say, how easy or hard it is to use. To rate this difficulty of use, we use a five-point scale, with 1 being very easy and 5 being very hard. The specific criteria we use to determine this difficulty of use should be described in our codebook, along with any other instructions for coding the study's variables. (For space reasons, the sample sheet contains columns only for 10 years' worth of utensils; if we were to conduct this project—and who wouldn't want to learn more about the history of kitchen utensils?—we'd need columns for each of the 50 items in our sample.)

Table 15.2. Sample Code Sheet for Study of Kitchen Utensil Popularity over Time

	1970	1971	1972	1973	1974	1975	1976
Utensil name							
Utensil purpose							
Price (in current dollars)							
Ease of use (1 being easy, 5 being hard)							
Other notes							

As our example shows, a code sheet can contain both qualitative and quantitative data. For instance, our *ease of use* row will report our difficulty rating for each utensil, a quantitative assessment. We will be able to analyze the data recorded in this row using statistical procedures of the sort outlined in **Chapter 14: Quantitative Data Analysis**—say, by calculating the mean value of "ease of use" for each of the five decades we are observing. We will be able to do the same thing with the data collected in the "price" row, which is also a quantitative measure. The final row of our example

code sheet will contain qualitative data: notes about our impressions of the utensils we are examining. For the data in this row, conducting open and focused coding (as described in **Chapter 11: Qualitative Data Analysis**) is an option. But regardless of whether we are analyzing qualitative or quantitative data, our goal will be the same: identifying patterns across our data.

Let's delve more deeply into what the coding of manifest content looks like, drawing from the example of an actual published paper. In her study of scouting manuals published by the Boy Scouts and Girl Scouts, Kathleen Denny (2011) sought to understand how these two organizations communicated different expectations about genders. As we noted, the measurement process should be carefully controlled for a quantitative content analysis, and Denny was very specific in how she conceptualized and operationalized a key question in her study: did the Boy Scout and Girl Scout manuals differ in the types of activities that scouts participated in? (Refer to **Chapter 7: Measuring the Social World** for a review of the **conceptualization** and **operationalization** stages.)

First, Denny defined an "activity" as a set of instructions that the scout must accomplish to be awarded a particular badge. She defined "participate" as attempts to fulfill the required instructions. In terms of how the manuals differed, Denny focused on a distinction she made between "others-oriented" and "self-oriented" activities, which she conceptualized as follows:

> I refer to activities explicitly instructed to take place with or for others as "others-oriented activities" and activities either explicitly instructed to take place individually or not explicitly instructed to take place in groups as "self-oriented activities." An example of a self-oriented activity is, "Draw a floor plan of your home," for the boys' Engineer badge . . . . An example of an others-oriented activity for the girls is, "In a troop, group, or with other girls, brainstorm a list of ways you can help the environment," for the Your Outdoor Surroundings badge. (Denny 2011:34)

Next, Denny decided on operational definitions for her concepts of "self-directed" and "others-directed" activities. Later in her article, she details how these variables were measured—again, crucial context for any content analysis:

> Activities coded as self-oriented are not necessarily required by the organization to be executed alone. Activities were coded as self-oriented if the activity description specifically indicated that it should be accomplished individually or implied individual action (e.g., reading) or if the activity instructions did not specifically call for the presence of others. For instance, the activity "Visit the newsroom of a newspaper" . . . was coded as a self-oriented activity because the instructions do not include others, whereas the activity "In a troop, group, or with other girls, brainstorm a list of ways you can help the environment" . . . was coded as others-oriented because the instructions expressly call for others to be present. (Denny 2011:44)

Denny reviewed 1,763 badge activities across the Boy Scout and Girl Scout handbooks. By using a structured coding process to analyze the handbooks' manifest content and count up the activity types they signaled, she was able to conclude that boys were offered more self-oriented activities, while girls were offered more others-oriented activities. (Since the time of Denny's study, the Boy Scouts have rebranded their flagship program as "Scouts BSA" and have started admitting girls, which may mean that the gender expectations in their handbooks have changed—yet another reminder that the context in which we conduct our analyses limits their scope.)

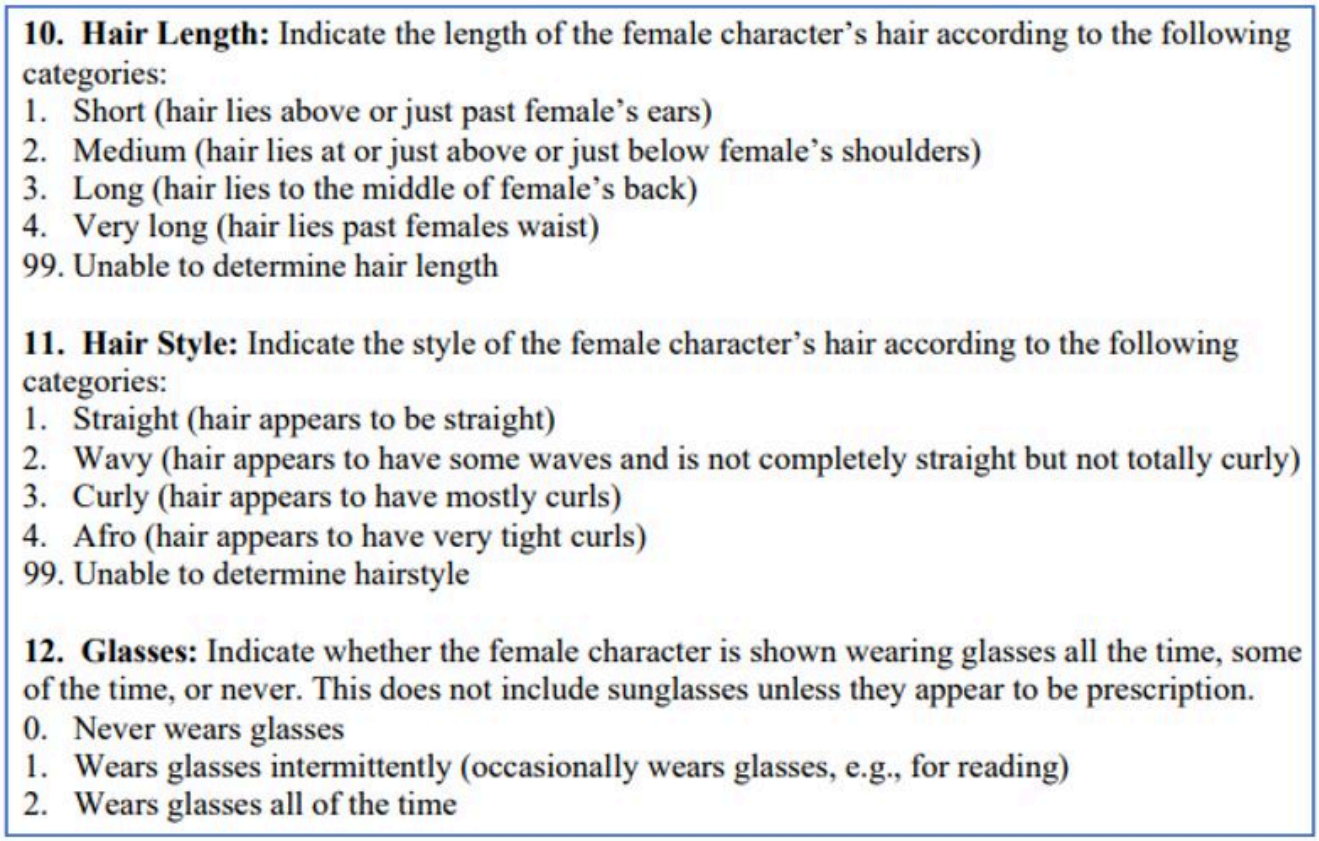
Denny was the only coder for her study, so she had to be especially scrupulous about consistently applying the measurement procedure outlined earlier. When two or more coders are involved, we have more checks in place to ensure the quality of the coding process—it is less likely to be the case that one coder's biases (however unconscious) skew the study's interpretations of the data. When multiple coders are involved, however, researchers must develop a detailed coding manual and actively train all the coders to ensure consistency.

Let's walk through how you should prepare to code content for a study, drawing as needed on examples from the content analysis used for the James Bond films (Neuendorf et al. 2010), which involved eight coders.

1. **Determine the variables to be measured in the analysis.** One of the research questions in the content analysis of James Bond films was, "Is the amount of aggression aimed at a female character predicted by her physical characteristics, role prominence, sexual activity, and aggressive predispositions?" (Neuendorf 2010:750). Some of the variables in this question that needed to be measured were:

    - If a person appears in the film, is that person a character?
    - If the person is a character, what is their gender?
    - If the person is a female character, how much aggression is aimed at her?
    - If the person is a female character, what is the nature of her physical characteristics?

2. **Determine the possible attributes of each variable.** Remember that **attributes** are the different categories for a given variable. In a content analysis, these are the options that coders can choose from. Specifically, they will code particular snippets of data as reflecting those attributes. For simple variables, this process of identifying attributes is pretty straightforward. In the Bond study, the "role prominence" variable—which indicated how major or minor a female character was in regards to events of the film overall—had just three attributes: minor, medium, or major. More complex **multidimensional** concepts, however, may require multiple variables to measure them. For example, the "physical characteristics" variable we just described needed to be measured using eight variables, including "hair color," "hair length," "glasses," "body type," and "level of attractiveness of physical appearance."
3. **Develop a coding manual providing precise instructions for coding.** The coding manual should describe all of the study's variables and give coders a complete understanding of how to make decisions during the coding process. This step is of paramount importance when multiple coders are used, which was the case in the Neuendorf study. (**Figure 15.2** shows actual instructions from their full coding manual.) Even with just one coder, however, creating a set of clear instructions will promote consistency throughout the coding process.
4. **Test the coding instructions on similar materials to revise your coding manual and train multiple coders.** Before you dive into your content analysis of your sources, it's a good idea to take materials that are somehow comparable to those sources and try out your variable definitions on that content. You can then refine your coding manual based on the outcome of that trial run. Testing the instructions is especially useful when you have multiple coders involved. That testing can serve as part of the training for coders, in addition to whatever informal or formal instruction you wish to give them concerning the coding manual. You can also use the test to check the level of agreement among coders. (In the section **Reliability and Validity in Content Analysis**, we'll discuss how you calculate the inter-coder reliability of the coding process for a given study.) For instance, the Neuendorf study was able to evaluate the level of agreement among its eight coders by using a Bond film (*Never Say Never Again*, 1983) that was not in the actual study list because it was based on an earlier Bond film (*Thunderball*, 1961). This pilot test prompted the researchers to make several changes to the coding manual and have the coders attend additional training sessions. They conducted their final training using the 1967 version of *Casino Royale*, which featured James Bond as a character but had not been included on the study list because it was a spy parody not produced by the same company that produced the other Bond films.
   For this study, the researchers were lucky to have two comparable films at their disposal that they could use to test their measurement procedures. If you find it hard to find a similar set of materials for such purposes, however, you can conduct a test on a source you intend to include in your study: do the trial run to refine your coding manual, delete the data, and then code the same film with the revised instructions.



10. Hair Length: Indicate the length of the female character's hair according to the following categories:
1. Short (hair lies above or just past female's ears)
2. Medium (hair lies at or just above or just below female's shoulders)
3. Long (hair lies to the middle of female's back)
4. Very long (hair lies past females waist)
99. Unable to determine hair length

11. Hair Style: Indicate the style of the female character's hair according to the following categories:
1. Straight (hair appears to be straight)
2. Wavy (hair appears to have some waves and is not completely straight but not totally curly)
3. Curly (hair appears to have mostly curls)
4. Afro (hair appears to have very tight curls)
99. Unable to determine hairstyle

12. Glasses: Indicate whether the female character is shown wearing glasses all the time, some of the time, or never. This does not include sunglasses unless they appear to be prescription.
0. Never wears glasses
1. Wears glasses intermittently (occasionally wears glasses, e.g., for reading)
2. Wears glasses all of the time

Figure 15.2. Coding Instructions from Neuendorf et al. (2010). Before they began their analysis of the portrayals of women in James Bond films, Neuendorf and her colleagues created a 13-page codebook to guide their team of eight coders. As shown in the excerpt, the coding manual provided detailed instructions about how to decide which of a variable's attributes (categories) were being reflected in the data being analyzed (Neuendorf et al. 2010:4–5). *Springer*

Coding in Qualitative Content Analysis

Coding in qualitative content analysis is typically an inductive process. That is, researchers do not start out by precisely specifying relevant variables and attributes and creating detailed coding procedures. Instead, they let their ideas regarding conceptualization and operationalization emerge from a careful reading and consideration of the text (see **Chapter 4: Research Questions** for a fuller discussion of the **inductive approach** to analysis). In this and other ways, coding in qualitative content analysis follows much the same procedures that are used for other qualitative research methods, such as ethnographic observation and in-depth interviews. That said, keep in mind the earlier distinction we made between quantitative content analysis, which focuses on manifest content, and qualitative content analysis, which focuses on latent content. Rather than just counting more obvious details, qualitative content analysis tends to delve more deeply into the data. The researchers immerse themselves in the text through careful reading, trying to get at its underlying meanings.

Sociologist Nikita Carney (2016) analyzed the Twitter debates that occurred after the police shot and killed two unarmed African American men, Michael Brown and Eric Garner, in 2014. In her paper she describes the straightforward qualitative data analysis process she followed:

> I decided to use Twitter's advanced search feature and take screenshots of selected results between December 3 and 7, 2014. The analysis process drew heavily from grounded theory [an inductive approach] in order to identify key themes. . . . I initially read through approximately 500 tweets from this time period to get a sense for the dialogue on Twitter at this moment in time. Based on this initial read-through, I loosely coded tweets based on whether they used "#BlackLivesMatter," "#AllLivesMatter," or both. I selected 100 tweets out of the initial sample of 500, consisting of approximately 30 to 35 tweets from each initial grouping that were representative of the larger sample. I conducted a close textual analysis on these 100 tweets, from which I developed more specific thematic groupings, including "call to action," "conflict over signs," and "shifting signs/discourse." (Carney 2016:188–9)

Just as different people who read the same book will not necessarily have the same interpretation of the text, closely studying latent content is necessarily a subjective process. Qualitative researchers recognize this and often disclose their personal stances toward the research, in a process known as reflexivity (discussed in **Chapter 9: Ethnography**). Carney's paper includes a "personal reflexive statement" to this effect:

> I closely followed news surrounding the deaths of Michael Brown and Eric Garner, among other victims of police violence, on mass media and social media. I also took to the streets with other activists and participated in acts of protest at my university on a daily basis. Rather than claiming to produce an "objective" analysis, I use my subjectivity to examine discourse as it unfolded on social media with the goal of better understanding the ways in which youth of color used technology to influence dominant discourse in the nation. (Carney 2016:181)

Although content analysis often focuses on either latent or manifest content, the two approaches can be combined in one study. For instance, we previously discussed how Kathleen Denny's content analysis of scouting handbooks examined manifest content—the descriptions of badge activities that could be categorized as "self-oriented" or "others-oriented"—and calculated which proportion of activities for the Boy Scouts and Girl Scouts fell into each category. But Denny also analyzed latent content in terms of how the handbooks portrayed gender. Based on this analysis, Denny (2011:27) argued that the girls were encouraged to become "up-to-date traditional women," whereas boys were urged to adopt "an assertive heteronormative masculinity." In her paper, Denny described the qualitative and inductive approach she took to arrive at this finding:

> Rather than code the texts for the presence or absence of any particular trait, I assessed them holistically, attuned to themes and patterns that emerged having to do with the attitude or approach endorsed by the texts. I performed textual analyses of the girls' and boys' handbooks' official statements as well as a focused comparison of a comparable pair of badges—the girls' Model Citizen badge and the boys' Citizen badge. I present findings from the comparison of the citizen badges because the nature of the activities offered in these badges is very similar, bringing gender differences into sharper relief. (Denny 2011:35)

Given the large amount of data that content analysis typically involves, it is often a good idea to conduct the coding using qualitative data analysis (QDA) software, which we described in **Chapter 11: Qualitative Data Analysis**. For instance, the content analysis mentioned earlier of undergraduate comments on a national survey (Chambers and Chiang 2012) used the NVivo software package to help researchers inductively generate codes and then consolidate, refine, and prioritize certain codes as they reviewed students' answers to the survey's open-ended questions.

Reliability and Validity in Content Analysis

As we noted earlier, having multiple coders can help address problems of subjectivity and bias that creep into a content analysis, but it also poses issues with **inter-coder reliability** (also known as **inter-rater reliability**, which we discussed in **Chapter 7: Measuring the Social World**). To ensure the reliability of their measures of different variables, projects that use multiple coders should discuss the degree to which different coders agreed upon how to code each variable in

the study. As we will describe, there are statistics that researchers can calculate to convey the level of agreement and disagreement among coders. At the very least, any project involving coding needs to say something about the ways the researchers ensured that their findings were not the result of an idiosyncratic or unreliable coding process. Even if one person acts as the sole coder, having one other person code a sample of the material will allow for at least some check of inter-coder reliability.

In the study by Neuendorf and her collaborators (2010) of women's portrayals in James Bond films, the researchers calculated two measures commonly used to assess inter-coder reliability: **multiple-coder kappa** and **percentage agreement among coders**. The corresponding statistics for each variable in their study are shown in **Figure 15.3**. "Percentage agreement" is the simplest measure of inter-coder reliability: in this study, it is the number of times that the eight coders involved in the study arrived at the same variable attribute for a particular observation (of one female character in one movie, for instance), divided by the total number of observations. The details of calculating kappa are beyond the scope of this textbook, but a kappa value of .40 or higher is generally considered to be acceptable.

Table 2 Intercoder reliabilities.

Variable	Multiple coder kappa	% Agreement	Lin's concordance (avg.)
Role	.66*	81	
Good/Bad	.33	71	
Race	.62*	89	
Hispanic identity	.47*	86	
Accent	.44*	70	
Hair color (5-cat)	.65*	76	
Hair Length	.58*	63	
Hair Style (3-cat)	.50*	76	
Physical Appearance (3-cat)	.48*	76	
Chronological Age (years)			.89**
Body Size (1–9)			.47**
Sexual Activity		79–100	.75**
Harm		88–100	.56**
Use of Weapons		88–100	.80**
Victim of Weapons		82–100	.78**
Kill Bond	.53*	88	
Consequences	.65*	80	

*Meets criterion of .40 recommended by Banerjee et al. (1999)
**$p<.01$

Figure 15.3. Calculating Inter-coder Reliability in Neuendorf et al. (2010). The content analysis by Neuendorf and her collaborators of women's portrayals in Bond films involved a team of eight coders. In their paper (Neuendorf et al. 2010:753), the researchers presented multi-coder kappa statistics that evaluated the degree of agreement among coders in coding certain variables (a kappa value of .4 is considered moderate, with higher values denoting higher agreement).

As shown in the figure, one variable—whether the female character was a "good" or "bad" person—did not meet the .4 threshold for the kappa statistic of reliability. This result isn't surprising, given that coding the variable required a broad evaluation of whether the female character was "good" (exhibited behaviors that furthered Bond's goals) or "bad" (exhibited behaviors that were at odds with Bond). In their paper, the researchers said they suspected that ambiguity may have arisen regarding females with minor or medium roles. They concluded that results related to this variable should be interpreted with caution. Note that the percentage agreement among coders was also low for the "good/bad" variable, but not as low as this reliability measure was another variable—whether the character had an accent—which highlights the fact that different inter-coder reliability measures can produce different results, and you may not want

to rely on just one. Furthermore, just evaluating inter-coder reliability after the coding is done is not the best strategy; instead, you should test the reliability of your coding procedures before implementation, which will give you the opportunity to improve upon them and possibly keep your reliability measures high across all variables.

As you might remember from **Chapter 7: Measuring the Social World**, the **validity** of a measure refers to whether the measurement procedure actually measures the concept it is intended to measure. Several approaches to assessing validity are discussed in that earlier chapter, but we will mention two that are highly relevant to content analysis. **Face validity** tells us whether it is *plausible* that our operational definition of a variable measures the concept it is intended to measure: "on the face of things," does our particular measure capture the essence of that concept? In *The Content Analysis Guidebook*, Neuendorf (2017:125)—the lead author of the Bond content analysis—points out that assessing a measure's face validity might seem "deceptively simple," but it is actually highly useful, requiring that "the researcher take a step back, so to speak, and examine the measures freshly and as objectively as possible." In fact, Neuendorf advises researchers to go one step further in ensuring face validity: "have others review the measures, with no introduction to the purpose of the study, and have them indicate what they think is being measured, a kind of back translation from operationalization to conceptualization."

If a measure has **content validity**, it is said to cover all of the domains or dimensions of a concept that it should cover. This is a critical consideration for any variable included in your content analysis, and achieving content validity may mean breaking up a complex variable into several measures, as we mentioned earlier. Consider the variable in the Bond content analysis that we said failed to meet the kappa threshold for inter-coder reliability: whether the female character was "good or bad." The coded attributes of this variable were as follows:

1. Starts off as good but turns bad.
2. Starts off as bad but turns good.
3. Is bad throughout the entire film.
4. Is good throughout the entire film.
5. Unable to determine.

We can see why there was ample disagreement in coding this variable, given the complexity of its underlying notion of "good/bad." For instance, what if a female character vacillated between good and bad during the course of the film? Should that option (or others) be added to cover all the possible dimensions of this underlying concept? These are the sorts of decisions you will need to make when operationalizing key concepts in a content analysis. You need to be able to defend your operational definitions as valid measures, keeping in mind that there is not just one "right" way to set up your variables.

Presenting Content Analysis Results

As we previously alluded to, quantitative content analysis creates numerical datasets that can be analyzed using any of the statistical techniques available to sociologists (see **Chapter 14: Quantitative Data Analysis** for an in-depth discussion). Although advanced multivariate techniques can certainly be applied, many content analyses published in academic articles just employ univariate and bivariate techniques like the ones we'll demonstrate now.

Earlier, we discussed a content analysis by Chambers and Chiang (2012) of undergraduate comments in response to open-ended survey questions. This analysis was initially qualitative: the researchers coded the undergraduate comments to inductively generate themes. However, the coding scheme they created allowed them to go back and label each of the 843 student comments as exemplifying one or more of the study's nine overarching code categories. **Figure 15.4** shows how the researchers were able to produce univariate statistics after quantifying aspects of their data. As shown in the frequency table, academic experience represented the largest percentage of total codes.

Table 3 Code category, ranking and frequency list of 2008

Ranking	Categories	Code counts	Percentage	Theme groups
1	Academic experience	520	38	Academic needs
2	Social	220	16	Non-academic needs
3	Campus	180	13	Non-academic needs
4	Level of satisfaction	163	12	Student satisfaction
5	Student services	125	9	Non-academic needs
6	Financial issues	90	7	Non-academic needs
7	Survey	36	3	Other concerns
8	Other	30	2	Other concerns
9	Ethnic concerns	8	1	Other concerns

Figure 15.4. Frequencies of Codes in Student Comments: A Content Analysis by Chambers and Chiang (2012). This table appeared in a content analysis of student comments in the National Survey of Student Engagement National Survey (Chambers and Chiang 2012:1117). In this study, researchers initially generated their codes through qualitative analysis. Then they went back and counted the appearance of one or more of these codes in each of the 843 student comments—thereby quantifying the prevalence of particular concerns across these comments. Note that the number of codes tallied in the table is larger than the number of students because students could address more than one issue in a comment. *Springer*

The frequency distributions calculated through quantitative content analysis can also be summarized in graphical form. **Figure 15.5** presents a chart from a content analysis by Francesca Cancian and Steven Gordon (1988). The researchers examined a sample of marital advice articles from U.S. women's magazines published between 1900 and 1979. For their analysis, they determined whether each article presented traditional or modern norms about marriage. According to their definitions, a "traditional" perspective held that romance was immature and women should prioritize their household roles; a "modern" perspective promoted passion and impulse and preferred that women take on roles outside the household. The researchers captured the degree of "traditional" versus "modern" content across their sources by using a composite measure that drew on various codes. The graph depicts changes over time in the percentage of articles advocating modern norms.

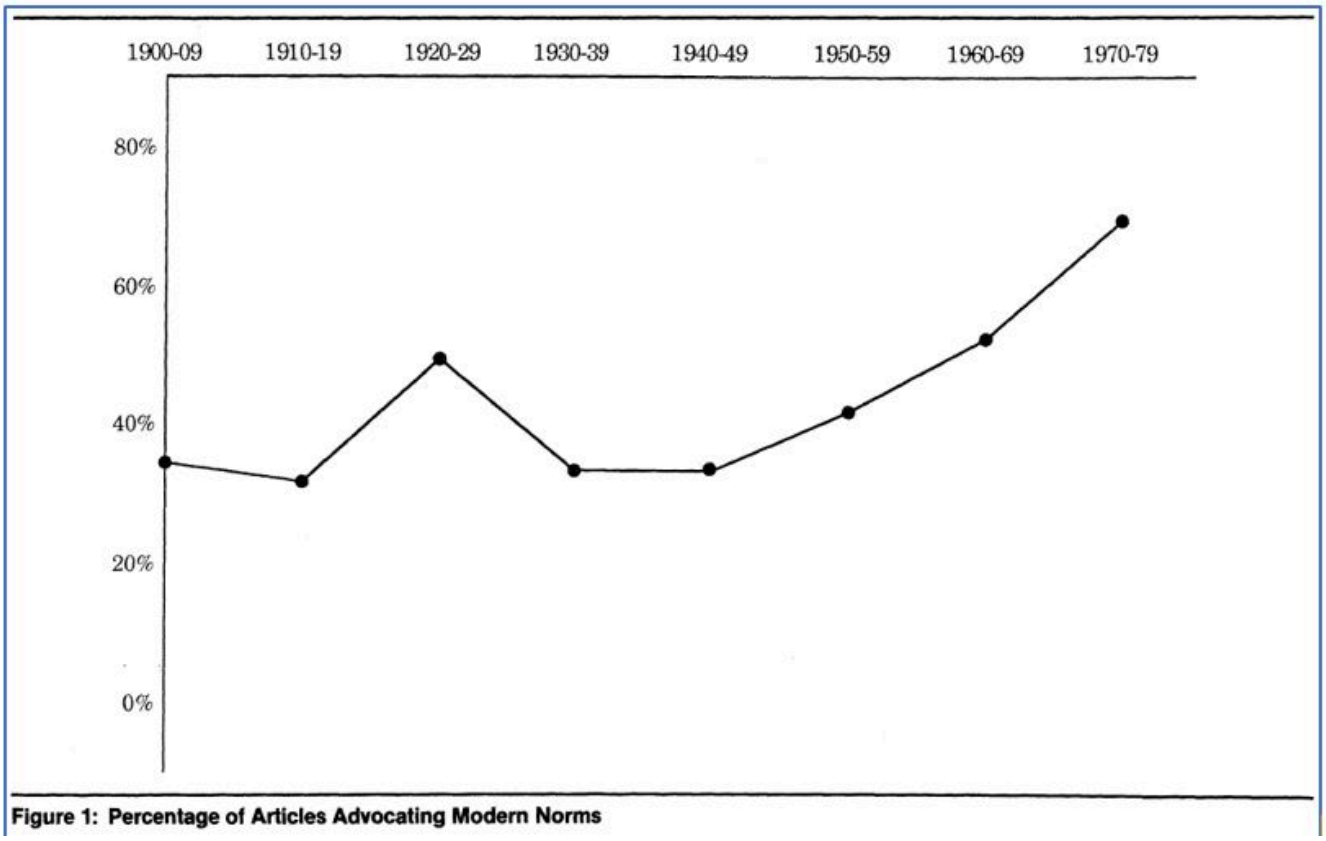


Figure 15.5. Views about Marriage Norms in U.S. Women's Magazines from 1900 to 1979: A Quantitative Content Analysis by Cancian and Gordon (1988). This content analysis of magazine articles providing marriage advice found that the share of articles endorsing modern norms about marriage fluctuated between 1900 and 1940 but then started rising dramatically (Cancian and Gordon 1988:327). *Sage*

As we discussed in **Chapter 14: Quantitative Data Analysis**, bivariate analysis allows researchers to examine the relationships between two variables. Usually, one is considered to be the independent variable, and the other is treated as the dependent variable that changes in response. In the analysis of scouting handbooks by Denny (2011) that we mentioned earlier, the independent variable was cultural understandings of gender (operationalized through Boy Scouts and Girl Scouts materials) and one of the dependent variables was activity type (coded as either "self-oriented" or "others-oriented"). Denny's analysis of these two variables is depicted in the graph in **Figure 15.6**.

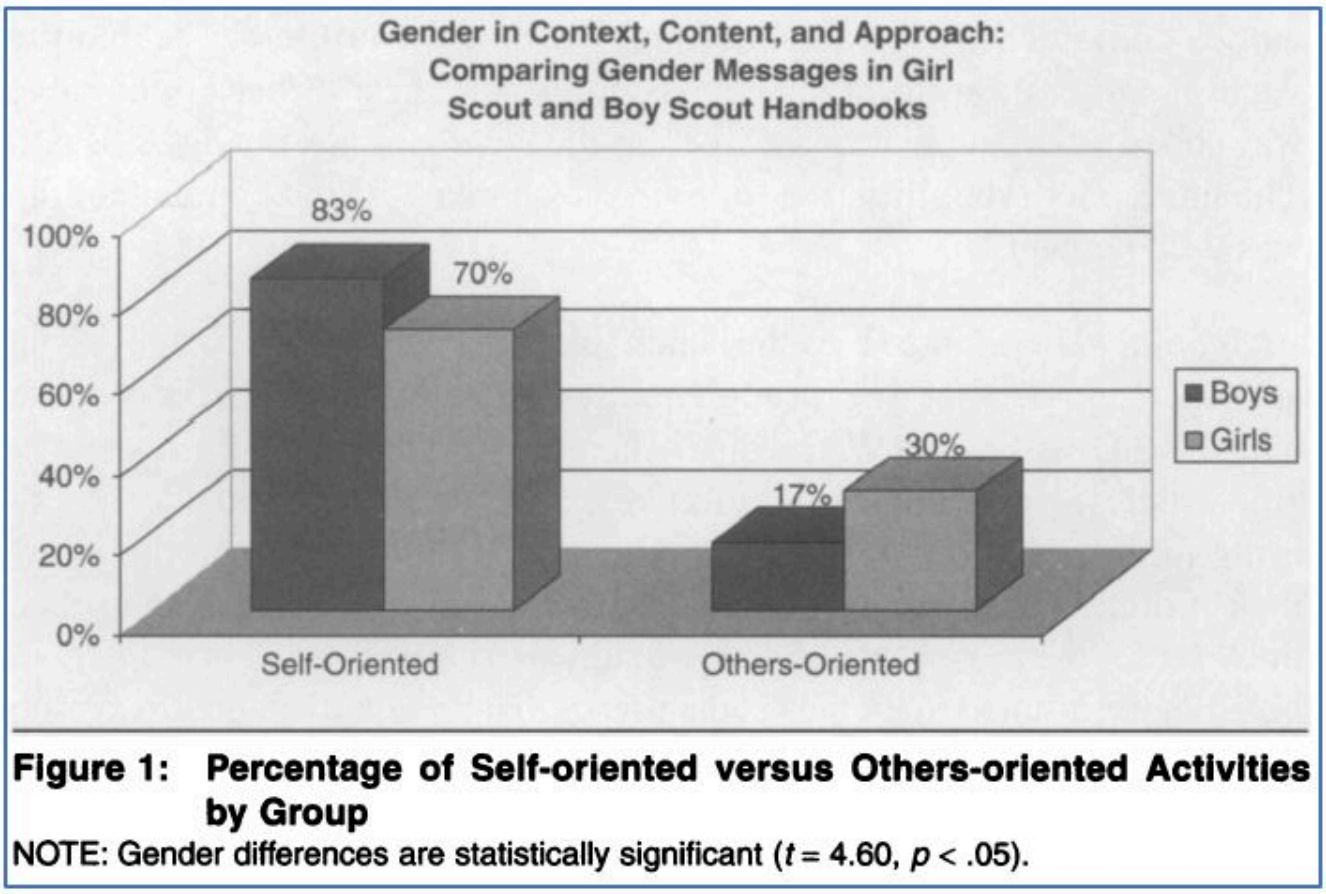


Figure 15.6. The Relationship between Gender and Scouting Activity Type: A Content Analysis by Denny (2011). A content analysis of scouting materials found that Girl Scout activities were more likely to be "others-oriented"—that is, designed to take place with or for others—than Boy Scout activities were (Denny 2011:36). As depicted in the chart, 30 percent of Girl Scouts activities fell into this category, compared with 17 percent of Boy Scout activities. *Sage*

A quantitative content analysis can also produce tables examining bivariate relationships. **Figure 15.7** presents a crosstabulation table created by Carol Auster and Claire Mansbach (2012) for their content analysis of toys marketed on the Disney Store website. Disney had classified 410 of its toys as being for boys and 208 as being for girls, with another 91 toys shown on both lists (which the researchers put in a third category of being for "both boys and girls"). The crosstabulation shows the relationship between the gender category of the toys (the study's independent variable) and the color palette used for the toys (the dependent variable).

Table 2 Color palette by the three-category classification of toys

Classification of toys[ab]	Color palette				
	Bold		Pastel		Total
	%	*n*	%	*n*	*n*
Boys Only	91.2	291	8.8	28	319
Both Boys & Girls	90.1	82	9.9	9	91
Girls Only	48.7	57	51.3	60	117

[a] Based on Disney's categorizations

[b] X^2 $(2, n=527)=108.2, p<.001$

Figure 15.7. The Relationship between the Gender Category and Color Palette of Toys Sold on the Disney Store Website: A Content Analysis by Auster and Mansbach (2012). As shown in this crosstabulation table, just half of Disney Store toys that were marketed as being for girls featured only bold colors, compared with roughly 90 percent of those for boys only and for both boys and girls (Auster and Mansbach 2012:382). *Springer*

When presenting the results of qualitative content analysis, sociologists may also break down numerical data in the ways we have just described. However, the focus is usually on interpreting the latent meanings of the text being analyzed. As you learned in **Chapter 11: Qualitative Data Analysis**, qualitative researchers often use quotes from their interviews to illustrate their participants' experiences and points of view. In a similar fashion, a qualitative content analysis will include quotes from the original sources to support the researchers' arguments. Consider the following excerpt from the content analysis of marital advice articles mentioned earlier:

> In this century's early decades, many articles explicitly expected the woman to do the "emotion work" in marriage. A 1932 article in *Ladies' Home Journal*, "And So You Are Married," told readers: "Well, whether your marriage is a success or failure depends upon you, little bride. . . . For marriage is not only a woman's vocation, it is her avocation. It is her meal ticket as well as her romance. . . . So it is not only good ethics but good business for a young wife deliberately to set about keeping her husband in love with her." (Cancian and Gordon 1988:322)

Through the use of quotations and paraphrases, qualitative researchers can provide a richer understanding of what the texts they are analyzing are actually communicating—in this example, how marriage advice articles imparted particular beliefs about marriage to their readers.

Strengths and Weaknesses of Content Analysis

Content analysis has a number of advantages over other research methods:

1. **The lack of contact with human participants avoids or minimizes some challenges that confront other methods.** Content analysis involves analyzing existing texts, rather than collecting data by direct contact with human participants. As a result, methodological issues like nonresponse bias tend to be less of a concern. Ethical scrutiny is less complex than it is for studies more directly involving human participants. That said, you should always remember that the content you are analyzing was created by human beings and is therefore subject to various biases related to reactivity, social desirability, and the many other ways that people's interests and perceptions shape what they communicate.
2. **Many materials are available at low or no cost.** For students in particular, content analysis is often the most convenient way of conducting original research, given that a wide variety of content sources are available online for free or via a university library's physical or digital collections. Refer to our lists of possible materials and example studies at the beginning of this chapter and in **Table 15.1** for inspiration about the types of sources you can draw upon.
3. **Content analysis can easily apply qualitative and quantitative methods to the study of materials.** As we've discussed, not only can content analysis examine texts in a quantitative or qualitative fashion, but it can easily incorporate both approaches. The coding can proceed deductively, with the coding categories created in advance, or inductively, with the categories emerging from researchers' careful classification of the raw data. And it can move between these approaches, as shown in previous examples—for instance, by starting with the inductive generation of codes and then going back to the data to quantify the existence of certain patterns based on those codes.
4. **Longitudinal analysis may be possible.** Researchers often use content analysis with historical materials, allowing them to provide insight into the evolution of social structures and processes over time. For instance, the content analysis described earlier of marital advice articles in magazines was able to cover a huge span of years—from 1900 to 1979—because of its tight but productive focus, and the result was a sweeping analysis of changes in emotional norms in marriage over decades (Cancian and Gordon 1988). (Note: we discuss historical-comparative research in more depth in the next section.)
5. **Content analysis has a high degree of flexibility.** Content analysis can be applied to a wide variety of materials and affords researchers substantial flexibility in the research process. Research methods that prove to be inadequate can be discarded and the materials can be reanalyzed using different techniques. Content analysis also offers flexibility in terms of when the research has to be conducted. Once accessed, the materials are typically available for study at any convenient time.
6. **A content analysis tends to be easy to replicate and update.** If researchers were systematic in selecting their sample and provided ample documentation regarding how they went about their coding, their content analysis can be easily replicated by others. This is especially true for quantitative content analysis, as it does not involve subjective interpretations of the data.

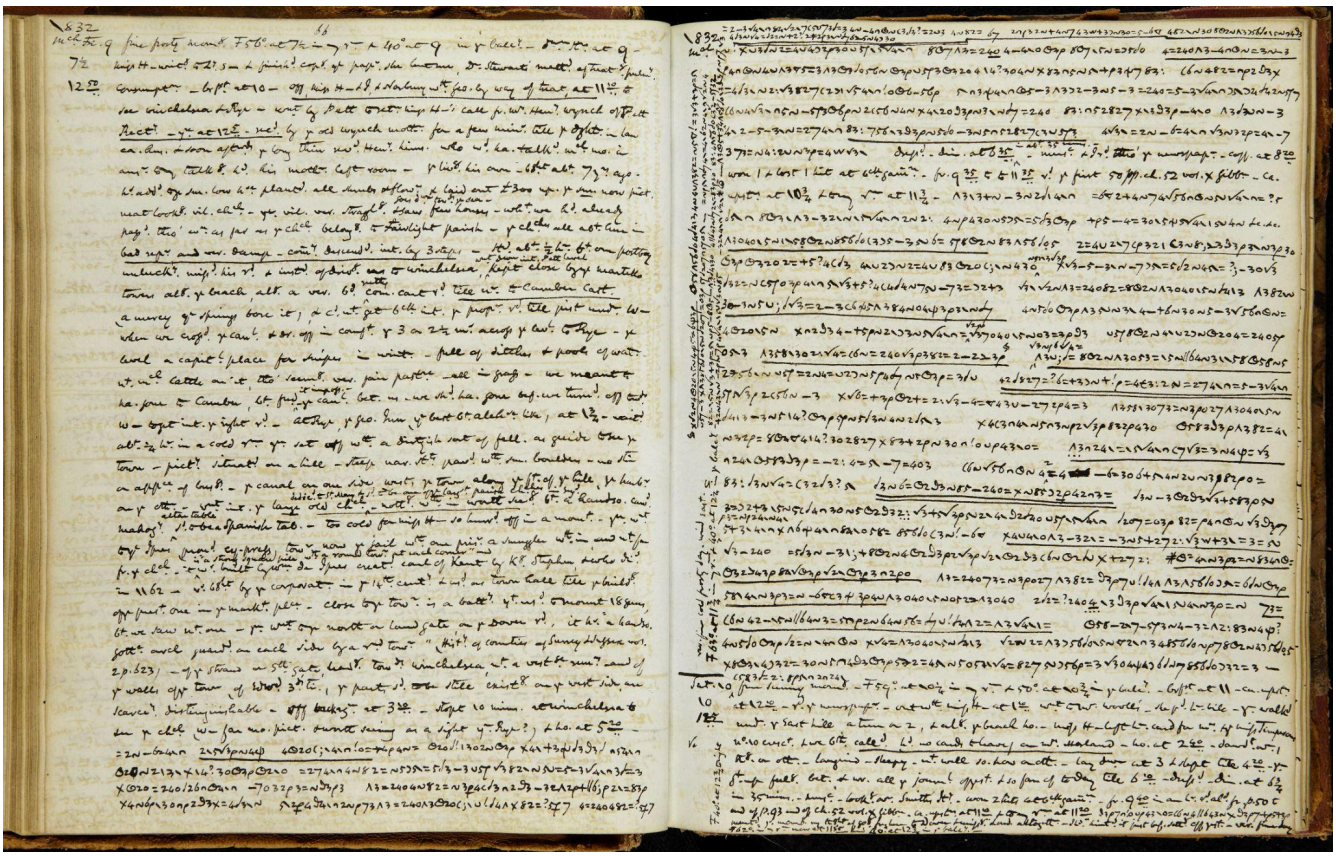


Figure 15.8. Challenges in Coding: Pages from Anne Lister's Diaries. Historians believe that Anne Lister (1791–1840) and her life partner Ann Walker (1803–1854) had the first (unofficial) lesbian marriage in the United Kingdom. Lister wrote in her diary daily and left behind 26 volumes with more than five million words, as well as an additional 14 travel diaries. In 2011, Lister's diaries were added to the UNESCO UK Memory of the World Register in recognition of their substantial cultural significance. However, the texts remain in a mostly untranscribed and undeciphered form, which means that it would currently be quite difficult to conduct a content analysis of them. *Calderdale Museums, via MyLearning*

Weaknesses of content analysis research include the following:

1. **Desired materials may not be in a format that can be studied.** Materials that the researcher would like to subject to content analysis may be hard to decipher. Anne Lister's extensive diaries provide a good illustration of this problem. Lister was a well-to-do British landowner in the early nineteenth century whom historians consider to be the first "modern lesbian." She left behind a massive set of diaries containing five million-plus words—the longest personal journal known in the English language. However, as shown in **Figure 15.8**, the handwriting in Lister's diary entries is exceedingly hard to read. About a sixth of the content in her diaries is written in code of her own devising, which she utilized to conceal details about her sexuality and intimate affairs (examples of the code are shown in the lower portions of the excerpted pages). Although the code decryption key has been available to scholars for many years, relatively little of the diary is available in a deciphered and digital form. As a result, a researcher would find this material impossible to use in a content analysis project unless they personally want to invest an enormous amount of time decoding and transcribing an appropriate sample of its pages.[1]
2. **Content analysis can be time-consuming.** Even if materials are readily available, the time and effort required to shift through them and identify relevant information may exceed what you as a researcher can devote to a project.

Even if a content analysis seems straightforward at the outset, closer inspection may reveal that the text is too complex or the materials too extensive for the study to be feasible. Take the State of the Union addresses we considered earlier as possible content sources. The longest written address is Jimmy Carter's 1981 address, which is 33,667 words long. The spoken addresses are considerably shorter, but can still be thousands of words, with the longest being Joe Biden's 2023 address of 9,216. A researcher who wanted to go through every written and spoken State of the Union address would need to read millions of words. As we have seen, researchers often analyze a sample of materials instead of all available materials, but this approach raises the issue of whether the sample is representative of the entire population of materials.

3. **Coding may focus on relatively superficial content.** Nowadays, the widespread transcription of print texts by Google Books and other services allows researchers to search entire decades' worth of publications using keywords. Specialized text analysis software can also be used to quickly digitize print materials as needed. However, these technologies may increase the temptation to focus on manifest content—words or phrases that are easy to count across a voluminous body of sources. This approach can make for a relatively superficial analysis, without any attempt to get at the deeper meaning of the texts.
4. **Coding instructions can be difficult to develop for complex concepts.** Recall the difficulty researchers encountered when they sought to measure whether the female characters featured in James Bond movies were "good or bad" (Neuendorf et al. 2010). This variable did not attain an acceptable level of inter-coder reliability, and part of the reason may have been the difficulty of determining the "goodness" or "badness" of a character when the woman did not play a major role in the film. As this example illustrates, the coding process is not always straightforward, and even a rigorous research project may be unable to produce and train people in coding instructions that avoid such ambiguity and disagreement.
5. **Reliability cannot always be assessed with certainty.** If we have only one researcher inductively coding the data for a qualitative content analysis, we will always wonder if another researcher—with a different outlook and biases—would have come to the same conclusions. Yet the breadth and complexity of the materials being reviewed may make it impossible to use multiple coders. For example, if we were conducting a qualitative analysis of Anne Lister's diary, we would need to recruit coders who had enough familiarity with her life and writing style to interpret the otherwise-impenetrable text. To its credit, quantitative content analysis can more easily draw upon multiple coders and calculate measures of inter-coder reliability. As we saw earlier, however, different measures can arrive at somewhat different results, and there is no surefire way of knowing whether the coding process was sufficiently reliable.

Key Takeaways

1. Content analysis focuses on the study of recorded communications. The materials that can be analyzed include actual written texts, such as newspapers or journal entries, as well as visual and auditory sources, such as television shows, advertisements, or movies.
2. Content analysis usually focuses on primary sources, though in some instances it may also involve

1. That said, images of all 7,000-plus pages of Lister's diary are available in the West Yorkshire Archive Service's online catalog, and transcriptions for its earliest entries (1806 to 1814) are now available in PDF form. An extensive decryption and transcribing project is now underway to make the entirety of the content easily accessible.

secondary sources.

3. Quantitative content analysis tends to count instances of manifest content, surface-level details about the sources. Qualitative content analysis tends to focus on the underlying meanings of the text, its latent content.
4. Code sheets are used to collect and organize data for content analysis.
5. Quantitative content analyses often present univariate and bivariate statistics in the form of tables and charts. Much like in-depth interviewing, qualitative content analyses tend to use direct or paraphrased quotes in order to describe meanings and explain relationships in the data.

Exercises

1. Identify a research question you could answer using content analysis. Now state a testable hypothesis having to do with your research question. Identify at least two potential sources of existing data you might analyze to answer your research question and test your hypothesis.
2. Create a code sheet for each of the two potential sources of data that you identified in the preceding exercise.

15.2. References

Auster, Carol J., and Claire S. Mansbach. 2012. "The Gender Marketing of Toys: An Analysis of Color and Type of Toy on the Disney Store Website." *Sex Roles* 67(7):375–88. doi: 10.1007/s11199-012-0177-8.

Borzekowski, Dina L. G., Summer Schenk, Jenny L. Wilson, and Rebecka Peebles. 2010. "E-Ana and e-Mia: A Content Analysis of Pro–Eating Disorder Web Sites." *American Journal of Public Health* 100(8):1526–34. doi: 10.2105/AJPH.2009.172700.

Boykoff, Jules, and Ben Carrington. 2020. "Sporting Dissent: Colin Kaepernick, NFL Activism, and Media Framing Contests." *International Review for the Sociology of Sport* 55(7):829–49. doi: 10.1177/1012690219861594.

Buysse, Jo Ann M., and Melissa Sheridan Embser-Herbert. 2004. "Constructions of Gender in Sport: An Analysis of Intercollegiate Media Guide Cover Photographs." *Gender and Society* 18(1):66–81. doi: 10.1177/0891243203257914.

Cancian, Francesca M., and Steven L. Gordon. 1988. "Changing Emotion Norms in Marriage: Love and Anger in U.S. Women's Magazines since 1900." *Gender and Society* 2(3):308–42.

Carney, Nikita. 2016. "All Lives Matter, but so Does Race: Black Lives Matter and the Evolving Role of Social Media." *Humanity and Society* 40(2):180–99. doi: 10.1177/0160597616643868.

Carr, Deborah, Elizabeth Heger Boyle, Benjamin Cornwell, Shelley Correll, Robert Crosnoe, Jeremy Freese, and Mary C. Waters. 2018. *The Art and Science of Social Research*. 2nd ed. New York: Norton.

Chambers, Tony, and Ching-Hsiao Chiang. 2012. "Understanding Undergraduate Students' Experience: A Content Analysis Using NSSE Open-Ended Comments as an Example." *Quality and Quantity* 46(4):1113–23. doi: 10.1007/s11135-011-9549-3.

Crichlow, Vaughn J., and Christopher Fulcher. 2017. "Black Men Down: An Assessment of Experts' Quotes on Deadly Encounters with Police." *Race and Social Problems* 9(3):171–80. doi: 10.1007/s12552-017-9197-x.

Denny, Kathleen E. 2011. "Gender in Context, Content, and Approach: Comparing Gender Messages in Girl Scout and Boy Scout Handbooks." *Gender and Society* 25(1):27–47. doi: 10.1177/0891243210390517.

Downs, Edward, and Stacy L. Smith. 2010. "Keeping Abreast of Hypersexuality: A Video Game Character Content Analysis." *Sex Roles* 62(11):721–33. doi: 10.1007/s11199-009-9637-1.

Ferree, Myra Marx, and Elaine J. Hall. 1990. "Visual Images of American Society: Gender and Race in Introductory Sociology Textbooks." *Gender and Society* 4(4):500–533.

Messner, Michael A., Margaret Carlisle Duncan, and Kerry Jensen. 1993. "Separating the Men from the Girls: The Gendered Language of Televised Sports." *Gender and Society* 7(1):121–37.

Neuendorf, Kimberly A. 2017. *The Content Analysis Guidebook*. 2nd ed. Los Angeles, CA: Sage.

Neuendorf, Kimberly A., Thomas D. Gore, Amy Dalessandro, Patricie Janstova, and Sharon Snyder-Suhy. 2010. "Shaken and Stirred: A Content Analysis of Women's Portrayals in James Bond Films." *Sex Roles* 62(11):747–61. doi: 10.1007/s11199-009-9644-2.

Park, Hee Sun, Hye Eun Lee, and Jeong An Song. 2005. " 'I Am Sorry to Send You SPAM': Cross-Cultural Differences in Use of Apologies in Email Advertising in Korea and the U.S." *Human Communication Research* 31(3):365–98. doi: 10.1111/j.1468-2958.2005.tb00876.x.

APPENDIX A: PRESENTING, WRITING, AND PUBLISHING



Throughout this textbook, we have described how to create a research question and answer it using sociological research methods. Once you've completed your analysis, though, your project is not over. In many ways, it is just beginning. Research that sits idle on your computer is not of use to anyone. Most sociologists hope their work will have relevance to others besides themselves: peers, policymakers, even the general public.

In this appendix, we'll talk about the process of preparing to share your work with others, which begins with deciding what to share and with whom to share it. Then we'll talk about the two main ways of sharing your work: (1) presenting it as a research talk and (2) publishing it. This material is designed to be read at any point. If you are taking a methods class or writing a thesis, you may want to read this appendix early on to understand how to structure the end product—that is, your research talk, proposal, or paper.

Opening chapter image credit: *Jessica Lewis/thepaintedsquare, via Pexels. Adapted by Bizhan Khodabandeh.*

A.1. Deciding What to Share and with Whom to Share It



Now that you're done with the data collection and analysis, you need to think about how to write up and present your findings. As you begin the process, think carefully about the audience you want to reach with your scholarship: while researchers frequently toil alone at their desks as they write up their results, they are always doing their work in conversation—explicitly or not—with other people. *RF_studio, via Pexels*

Learning Objectives

1. Understand the importance of being open about your study's strengths and weaknesses.
2. Describe how the audience you choose to target will shape the approach you take to sharing your research findings.

At a fundamental level, research is always a public activity. While the work may be conducted by a lone individual in a private setting, the knowledge gained from that work needs to be shared to truly be "science"—which, as we've emphasized throughout the book, adds to a conversation among scientists, with each new study building on previous work. Science is a dialogue, not a monologue, and so it's critical to think about how best to dialogue with audiences about your work—and which audiences would benefit the most from hearing about it.

When you are preparing to share your work with others, you should make sure that you know the answers to the following questions:

1. Why did I conduct this research?
2. How did I conduct this research?
3. What conclusions can I reasonably draw from this research?
4. Knowing what I know now, what would I do differently?
5. How could this research be improved?

First, understanding why you conducted your research will help you be honest—with yourself and your readers—about your own personal interest, investment, or bias with respect to the work. Being able to clearly communicate how you conducted your research is also important. Describing your methodology in detail helps convince other people that your study was conducted competently and ethically. It provides a blueprint that other researchers could possibly follow if they wanted to replicate or build upon your work. With this in mind, you want to be thorough and completely transparent in detailing your data collection methods, sample and sampling strategy, and data analysis procedures.

Next, you want to think about the major strengths of your work. Looking back at it, you should have one or a few key findings or arguments that you wish to make based on your research. You'll want to prioritize the takeaways in whatever presentation or publication emerges from your work. Don't just give people a long list of interesting but random findings from your research. Instead, say explicitly what your most important points are, and make sure the rest of your results can be categorized as subpoints under those overarching takeaways.

The last two questions—*What would I do differently*? and *How could this research be improved*?—are meant to make you think about potential weaknesses in your work and how future research might build on what you've done. Given that sociology is a science that seeks to arrive at the most accurate understanding of social processes, it's crucial that we sociologists be upfront about all aspects of our research, including its limitations. Doing so helps ensure that others will be able to understand, expand on, and effectively critique our work. We need to be open to questioning and scrutiny if the science is to advance. In this important sense, putting forward scientific findings is very different from arguing a legal case: we're not really trying to win an argument, but better understand a phenomenon—which our critics will help us to do.

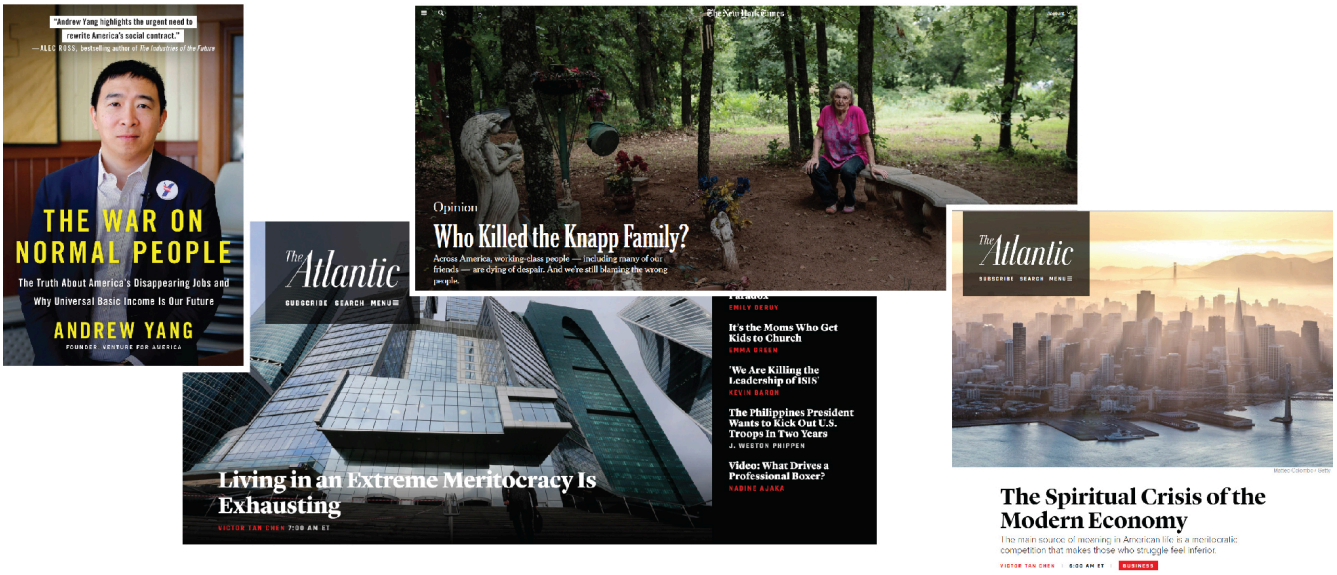
Remember that all scientific studies contain limitations. As we've emphasized throughout this book, research design requires you to make decisions that often have tradeoffs—between breadth and depth, generalizability and context, and so on. Furthermore, we human researchers—you included—aren't perfect. We conduct our research within the constraints of limited resources and time. As a result, there are bound to be less-than-ideal aspects to any research study, especially those done by novice researchers. Yet students often fret about talking about the flaws in their work—as if an examiner were evaluating them, red pen in hand, marking them off for every little mistake. While the occasional pesky journal reviewer will be like that, most readers of your work will be far more generous, knowing just how complex and challenging an undertaking an original research study is. They will appreciate your candor in discussing the inevitable problems in your work. With this in mind, avoid the temptation to overstate your study's findings or gloss over its shortcomings. If anything, overly perfect studies raise flags, given how messy and complex the study of social life always is.

Once you know what parts of your research you want to share, you should consider these three questions:

1. For whom did I conduct this research?
2. Who will read this?
3. Who will be affected by this?

Answering these questions will help you articulate who the **audience** and the major **stakeholders** are in your research. By "audience," we mean the people who will receive the scientific information you are sharing. If you are conducting research for a class project, your main audience will probably be your professor. Perhaps you'll also share your work with other students in the class. If you are conducting professional research, your audience will typically be other social scientists—the people who will read the academic journals or books you publish, or attend the academic talks you give. Researchers who study similar topics will probably be interested in your work. As we discussed in **Chapter 2: Using Sociology in Everyday Life**, these may be other sociologists, or they may be researchers from other disciplines who focus on your area of inquiry. They can include academics, government officials, experts at think tanks and other research organizations, or practitioners in fields like business, medicine, or social work who like to stay abreast of social scientific findings. Other potential audiences might be news reporters, *pundits* (experts with a public presence), and media representatives more broadly. Especially informed and inquisitive members of the public can also be a potential audience for compelling sociological work—though you will need to be doubly sure your work is written in a way that's accessible to them (more about that later).

We tend to think about audiences in a one-way fashion given that we share our research findings and do not expect them to weigh in. Stakeholders, on the other hand, are specific individuals or groups who are directly affected by your research and thus have a *stake* in those findings. If you conducted research for a job, your employer would be a stakeholder; if you received a grant to do the work, the foundation or government agency that funded you would be a stakeholder. As described in **Chapter 2: Using Sociology in Everyday Life**, other stakeholders might include **policymakers** who oversee laws or regulations relating to your research topic, or organizations that do business, advocacy, or other work in that area. People who share something in common with your research participants—such as members of the community where you conducted research, or members of the same social group you studied—might very well be interested in what you've discovered in your research, too, particularly if you engaged in the sorts of **community-engaged research** or **participatory action research** we described earlier. If you want to maximize the public impact of your scholarship, you will want to publish your work in a form that is easy for these stakeholders to understand—in other words, produce a version of your study where they are part of the intended audience.



Following the publication of his scholarly book *Cut Loose: Jobless and Hopeless in an Unfair Economy* (2015), Victor wrote a number of op-eds in the *Atlantic*, and his work was discussed in a book by Democratic presidential candidate Andrew Yang and in an article and book by columnist Nicholas Kristof and writer Sheryl WuDunn of the *New York Times*. *Hachette* / *Atlantic* / *New York Times* / *Atlantic*

Although you would never alter your actual findings for different audiences, understanding who your audience is will help you frame your research in a way that is most meaningful to that audience. For example, Victor has shared findings from his ethnographic study of autoworkers' experiences of long-term unemployment in the United States and Canada. He has published his work for a variety of academic and nonacademic audiences. The study was the focus of his book *Cut Loose*, published by the University of California Press, an academic press that utilizes peer review. The book was reviewed in sociology journals like *Contemporary Sociology* (the American Sociological Association's journal of book reviews) and the *American Journal of Sociology*. He published essays expanding on the material in his book in the *Atlantic* magazine and other news outlets. His work was also discussed in an article in the *New York Times* and a book by *New York Times* staff writers. Finally, he was interviewed by journalists and radio programs about the book.

Victor shared with all of these audiences the key findings from his research, but *how* he presented these findings, and the *level of detail* he shared about his research, varied by audience, as described in the sidebar **Writing Op-eds** later in the chapter. Generally speaking, **lay audiences** (those who aren't scholars) do not need as much background information about your methods of data collection or analysis. They won't be as interested as other scholars will be in evaluating how well you did your study. Lay readers, however, will want you to make clear what the broader significance of your findings are—the answer to the "So what?" question we always need to ask ourselves as researchers. They also tend to appreciate rich narratives and judiciously chosen statistics that crystallize your study's major themes.

Key Takeaways

1. When you are ready to write up your study, you should be able to summarize its key findings, prioritizing the most important results.
2. You need to be thorough and candid in describing any limitations of your research, which allows your work to be fairly evaluated and thus helps social science to progress.
3. The characteristics of the audience will shape which details about the study are emphasized, with lay audiences typically looking for less in the way of technical information about how the research was conducted.

Exercise

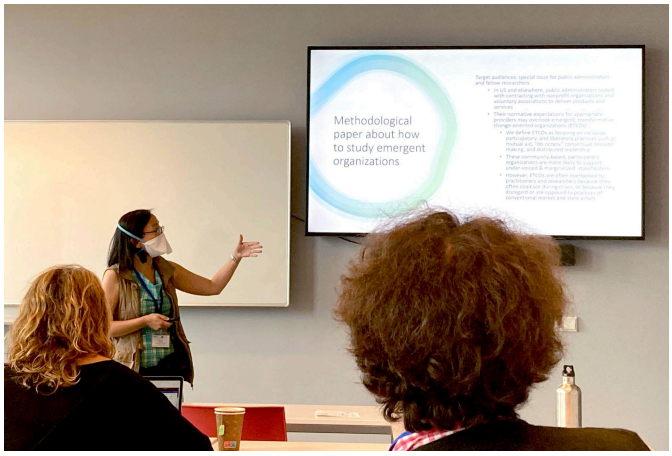
First, read the scholarly article "Gender, Health Behavior, and Intimate Relationships: Lesbian, Gay, and Straight Contexts" by sociologists Corinne Reczek and Debra Umberson (2012). What evidence can you find that might indicate that the authors gave some thought to the questions outlined in this section? Next, read this other article that Umberson wrote for the news site *Huffington Post* (2015) with a lay audience in mind. How did Umberson approach these academic and nonacademic articles in different ways, in terms of her writing style and choice of details to include?

A.2. Presenting Your Research

Learning Objectives

1. Understand best practices for giving formal presentations of research studies.
2. Describe the different approaches taken in roundtable and poster presentations.

Presenting your work not only allows you to disseminate your research findings, but it also is a tried-and-true method to improve your study-in-progress. Before submitting their research for publication, professional sociologists usually make presentations to peers about their study (or at least a portion of their work) so that they can test their ideas and get helpful feedback. If you are serious about research, you should get in the habit of presenting your work to audiences early on—well before publication—so that your work can benefit from other people's guidance. Just going through the process of summarizing your study for an audience unfamiliar with it is sure to sharpen the way you think about it.



At academic conferences, sociologists often present their papers as part of panels featuring three or more scholars. Typically, each presenter has 10 to 20 minutes to talk about their study's approach and key findings. At most conferences, students (including undergraduates) are welcome to present on panels, and their work is not distinguished as "student" work: if you do sociological research, you are a sociologist, and your work should stand on its own.

You can start by presenting your work to a small group—say, a fellow student or trusted colleagues—but eventually you want to move to an audience of less partial observers. Sociologists are sometimes invited to give **research talks** (presentations about their work) by academic departments, think tanks, or other interested organizations. The most common way that sociologists present their research, however, is by doing talks at **professional conferences.** Professional conferences can range from small-scale events involving a dozen researchers who convene for an afternoon, to large-scale events involving thousands of researchers who meet for several days. Professional associations (described in **Chapter 2: Using Sociology in Everyday Life**) typically put on these conferences on an annual basis.

There is a formal process for securing speaking spots at conferences. Usually, the organizers require you to submit an **abstract** of your research paper months in advance; some conferences may ask for extended abstracts (several pages) or even full papers, though this is rare in sociology. (Note that scholars use the term **paper** when talking about a write-up of a study that a researcher has conducted. A paper may be either published or unpublished—when it is an unpublished draft, we call it a **manuscript.**). The organizers review these submissions and decide which to accept. The peer-review process here is typically not as rigorous as it is for manuscripts submitted to an academic journal, and smaller conferences accept the vast majority of submissions. That said, your paper or proposal may get rejected, and you should bear in mind that a conference rejection isn't always a reflection of the quality of your work; it may be

because your topic doesn't fit with what the conference organizers are looking for. Talks are often grouped together into **panels** of three to four (sometime five) presentations that share an overarching theme. A professional association's annual meeting may also be devoted to a particular theme, meaning that any papers on that topic will be prioritized.

When preparing a formal talk, it is important to know well in advance how long your presentation is expected to last and whether any visual aids such as slides are expected. For conference panels, the typical talk lasts between 12 and 20 minutes. This may sound long, but you'll be amazed by how quickly the time flies. Once researchers get into the groove of talking about something as near and dear to them as their research, they often become so engrossed—and enamored with the sound of their own voice—they forget the clock. Many an unfortunate researcher has looked up to see their allotted time has expired before they have even gotten to their study's methodology. To avoid this, repeatedly practice your presentation in advance—and time yourself. A good rule of thumb is that every slide in your presentation will take one minute to get through (assuming you don't cram too much information into each slide!).

Another common mistake in research talks is taking too long to set up the study. Audience members will usually be more interested in your study's original findings than in hearing you recite a long list of previous studies it builds upon. While in a paper you need to discuss the literature at length (see **Chapter 5: Research Design**), in a presentation you should highlight *your* work.

When you are giving a research talk, never read your paper verbatim. In some fields, this is acceptable, but not in most traditions of sociology. Nothing will bore an audience to tears more quickly than a sociologist reading their **literature review** aloud word for word. In your talk, highlight only the key points of your study. As we describe in the next section, these details will generally include your research question, your methodological approach, your major findings, and a few final takeaways (generally, the theoretical implications of your work).

At academic conferences, panels almost always include time for audience members to raise questions or make comments, usually after all the session's talks have been given. These Q&As can be the highlight (or lowlight) of a given conference. In fact, some disciplines are notorious for the vicious questions and comments that audience members will lob at conference presenters—particularly if the scholars in question come from an opposing school of thought on a particular topic. Fortunately, sociologists tend to be more decorous during presentations, and unlike economists (renowned as the most dismal of social scientists), they tend not to repeatedly interrupt speakers during their talks. That said, expect audience members to comment on any issues with your research design, data, or logic—and be prepared with good answers (see the sidebar **Deeper Dive: Tips for Presentation Slides**).

Regardless of how friendly or unfriendly your audience's questions are, learn to respond to them graciously. It's polite not just to answer, but to note when you've been asked a good question, or to thank the questioner for raising a noteworthy issue. At the same time, be genuine and honest, and don't be afraid to say when you don't know something (you might even say you'll look into it and get back to the inquirer by email). At the end of any response you give, check in with the person who asked the question to make sure you truly answered their question.

If you are in the audience for a research talk, you shouldn't hesitate to ask a question or make a comment during the Q&A. Presenters want feedback. Speaking up gets easier with practice, and it is a good way to connect with presenters or audience members, who you can then approach after the talk. Of course, always be respectful. Given that you are not a curmudgeonly economist, don't interrupt the presenter during their talk unless you want to clarify something quickly. Learn to be diplomatic and constructive in your criticism or probing questions. Sociologists pride themselves on having rigorous but cordial debates at academic meetings, and hopefully you'll be able to find and cultivate a thoughtful intellectual community of fellow scholars at the events you attend.



Larger conferences hold *roundtable* sessions, where scholars are assigned to a group of authors with similar papers and then meet up to present and discuss their research. These sessions are not promoted by conferences as much as panel sessions, which means that fewer people are in attendance but also that the authors may get more substantial feedback from those who are there. *Gender Spectrum Collection, via Vice Media*

Larger conferences may offer other options for presenting your work in addition to panels. In less formal **roundtable** presentations, the aim is usually to help stimulate a conversation about a topic. The time you are given to present may be slightly shorter than in a formal presentation, and you'll also be expected to participate in the conversation that follows all presenters' talks. Because fewer people typically are in attendance, a roundtable is an excellent place to workshop your research when it's in its earlier stages—for instance, after you've conducted some pilot interviews for a longer qualitative study. You'll be able to talk through your initial findings and get some ideas about where to take the study next. In that more intimate setting, you'll probably receive more substantial feedback while also getting a preview of the objections reviewers may raise about your conclusions or your approach to the study.

In a **poster presentation**, you visually represent your work, putting together a printed-out summary of the study—much like a science fair board—that concisely describes its essential parts. During the scheduled poster session at a conference, other scholars will browse the displays, listen to you give your spiel about your research, and ask questions. Poster sessions, like roundtables, can be quite helpful at the early stages of a research project because they are designed to encourage the audience—here, people walking by—to engage you in conversation about your research. Just as you wouldn't read a paper verbatim in a talk, avoid just pasting sentences from your written report onto your poster. Instead, think about how to tell the "story" of your work in charts, tables, and images. Bullet points are fine so long as the poster

isn't so wordy that it would be difficult for someone leisurely walking by to grasp your major argument and findings. Don't try to include every detail about your work on your poster. The point is to share highlights and then converse with your audience to get their feedback, hear their questions, and provide additional context.

Sociology Professional Conferences



Following are links to the websites for a number of international, national, and regional sociology associations (some of which are mentioned in **Chapter 2: Using Sociology in Everyday Life**). All of these associations hold annual research conferences where students can present, and some sponsor panels or poster sessions expressly for the work of students, including undergraduate and master's students. Many sponsor their own journals, as noted, and some have job listings advertising academic or practitioner positions relevant to their members. (For more details about these organizations, visit this page.)

National and International Associations

- American Sociological Association: publishes a number of journals, including *American Sociological Review*
- Canadian Sociological Association: publishes the *Canadian Review of Sociology*
- International Sociological Association: publishes journals such as *International Sociology*

Regional Associations

- Eastern Sociological Society (ESS): represents East Coast sociologists; publishes *Sociological Forum*
- Mid-South Sociological Association: publishes *Sociological Spectrum*
- Midwest Sociological Society (MSS): publishes the *Sociological Quarterly*
- North Central Sociological Association (NCSA): publishes *Sociological Focus*
- Pacific Sociological Association: represents West Coast sociologists; publishes *Sociological Perspectives*
- Southern Sociological Society (SSS)

Group-Focused Associations

- Association of Black Sociologists: advocates for black sociologists and black people domestically and globally; publishes *Issues in Race and Society*
- Sociologists for Women in Society: promotes feminist social science and feminist leadership; publishes *Gender & Society*

Issue-Focused Associations

- Association for Applied and Clinical Sociology (AACS): represents applied, clinical, and public sociologists and sociological practitioners; publishes the *Journal of Applied Social Science*
- Rural Sociological Society: focuses on rural sociology; publishes *Rural Sociology*
- Society for the Study of Social Problems (SSSP): devoted to social research to further social justice and activism; publishes *Social Problems*

Interdisciplinary Professional Associations

Here is a list of some notable interdisciplinary associations that include sociologists among their members:

- Academy of Management (AOM): organizational studies; includes many scholars from business schools, and publishes several journals, including the *Academy of Management Journal*
- American Studies Association: U.S. culture and history; includes historians and literary scholars, and publishes *American Quarterly*
- Association for Public Policy Analysis and Management (APPAM): public policy and management; involves policymakers and practitioners, and publishes the *Journal of Policy Analysis and Management*
- Association for Research on Nonprofit Organizations and Voluntary Action (ARNOVA): nonprofit and philanthropic research; publishes *Nonprofit and Voluntary Sector Quarterly*
- Labor and Employment Relations Association (LERA): labor and employment studies; includes scholars and professionals from labor unions, human resources departments, and employment law, and publishes *LERA Annual Proceedings*
- Population Association of America (PAA): demography and related research; publishes *Demography*
- Society for the Advancement of Socio-Economics (SASE): economic sociology and political economy; publishes *Socio-Economic Review*
- Urban Affairs Association (UAA): represents urban scholars, researchers, and public service professionals; publishes the *Journal of Urban Affairs* and the *Journal of Race, Ethnicity and the City*
- Working-Class Studies Association (WCSA): research on working-class life and cultures; publishes the *Journal of Working-Class Studies*

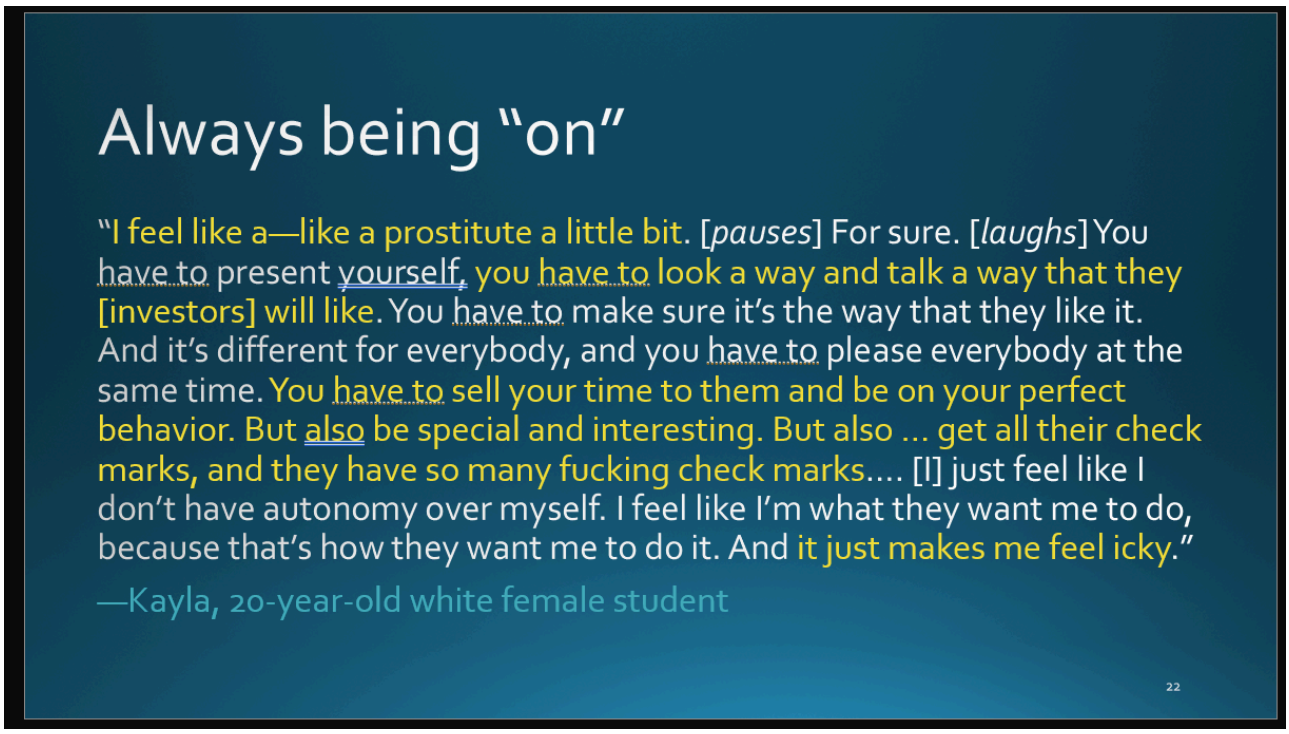
The Parts of a Research Talk

In sociology, a research talk typically follows this sequence. If you're using slides, the headings of your slides will use these or similar terms:

1. **Introduction:** The introduction provides general background, perhaps mentioning some relevant trends or statistics and otherwise making the case for the overall importance of the topic being studied. For shorter presentations, you may just skip to the next section. For longer presentations, it's often a good idea to start off with an outline of the sections of your presentation and then refer back to that outline as you move through each section (the **signposting** we talk about in the sidebar **Deeper Dive: Tips for Presentation Slides**). If you have time, it's a good idea to briefly identify the **research question** and **research problem** upfront**,** which you will return to

repeatedly in your presentation.

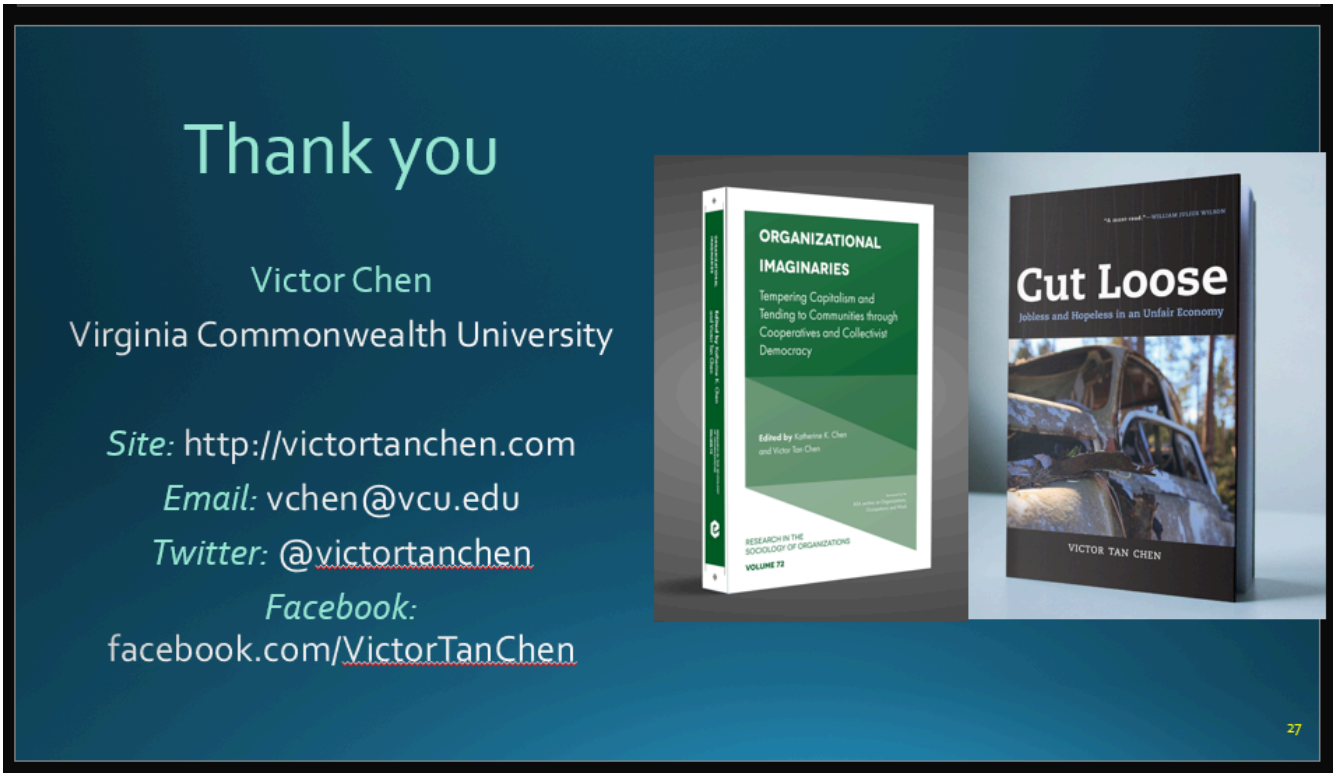
2. **Literature Review *or* Background:** As noted earlier, the literature review is very brief in a typical research talk. That said, you will want to mention the key studies that your work is responding to and building upon. *Don't* just list the studies; instead, your bullet points should refer to the specific arguments made in those past studies that are directly relevant to your own study. At the end of your bullet points, use in-text parenthetical ASA cites (see the section **Citing Your Sources**) to refer to any relevant studies—especially if you have several you want to mention at once—but don't bog down the slides with full citations. You also don't want to spend too much time on a single study unless the explicit purpose of your own research is to critique or build on that work. Instead, summarize the literature's overall argument in that area and include multiple cites as examples.
3. **Research Problem *or* Statement of the Problem *or* Puzzle:** If you've done your literature review slide correctly, you should have set yourself up well for this slide, which discusses the research problem—the gap in the literature that your study is addressing, or the puzzle in that literature it is resolving. (See **Chapter 5: Research Design** to learn how to identify and present your research problem or puzzle.) You should also state or restate the research question that arises from this research problem, perhaps on its own slide.
4. **Methods:** This should be a highly condensed version of the methods section you have in your written scholarly report. Start off by stating what the key methods or sources of data were—for example, in-depth interviews, participant observations, survey data, and experimental data. Make sure to mention the exact number of people interviewed or surveyed, the exact number of events attended, the exact date range during which the fieldwork or surveys were conducted, and so on—academics always want specifics. You may want a separate slide describing the **sample** (usually for quantitative studies) or *case selection* (usually for qualitative studies). In this slide, you should provide details like how representative the sample was (whether it was a probability or purposive sample, for example), how you decided on the populations or sites to focus on, and what your process for recruiting respondents was. For secondary data, be sure to mention the name of the dataset and any particular modules you used. Later in this section, describe the procedures you used for data analysis. For qualitative studies, you might talk about the theoretical approach that informed your data collection strategy (e.g., grounded theory, extended case method) and the coding process you engaged in (e.g., whether you used software packages like NVivo or Atlas.ti, how you generated codes from the data, whether you had multiple people code your data, and how you resolved any discrepancies in coding). For quantitative studies, you can talk about your choices of statistical procedures (e.g., logistical regression, hierarchical modeling) and statistical software packages (e.g., SPSS, SAS). Quantitative studies in particular will also have a set of slides devoted to **measures** or variables that break down the dependent variables (which usually go first), the independent variables of interest, control variables, and so on. Here, you will want to state explicitly how each relevant concept was measured. Include the verbatim question text for interview or survey questions that are especially important to your study, so that your audience knows exactly how you operationalized the underlying concept. Mention the scales or units for each measure.



If you are presenting qualitative work, you should always provide concrete details from at least a few of your cases—such as choice quotes or observations—to give your audience a sense of your actual data. Just summarizing the overall findings will make your research talk boring and unconvincing; show people what you actually did in the field. That said, when showcasing data from your in-depth interviews or fieldnotes in a talk, avoid quoting long passages, which can drone on. One approach is to put the entire quote on the slide but read only key portions of it (flagged in boldface or another text color, such as the yellow-colored text in this example). Another strategy is to read a (short) quote aloud but provide only a paraphrased summary of that quote on your slide, which means you won't distract the audience with a wall of text. Make sure to include basic demographic details about any interview respondents on the slide.

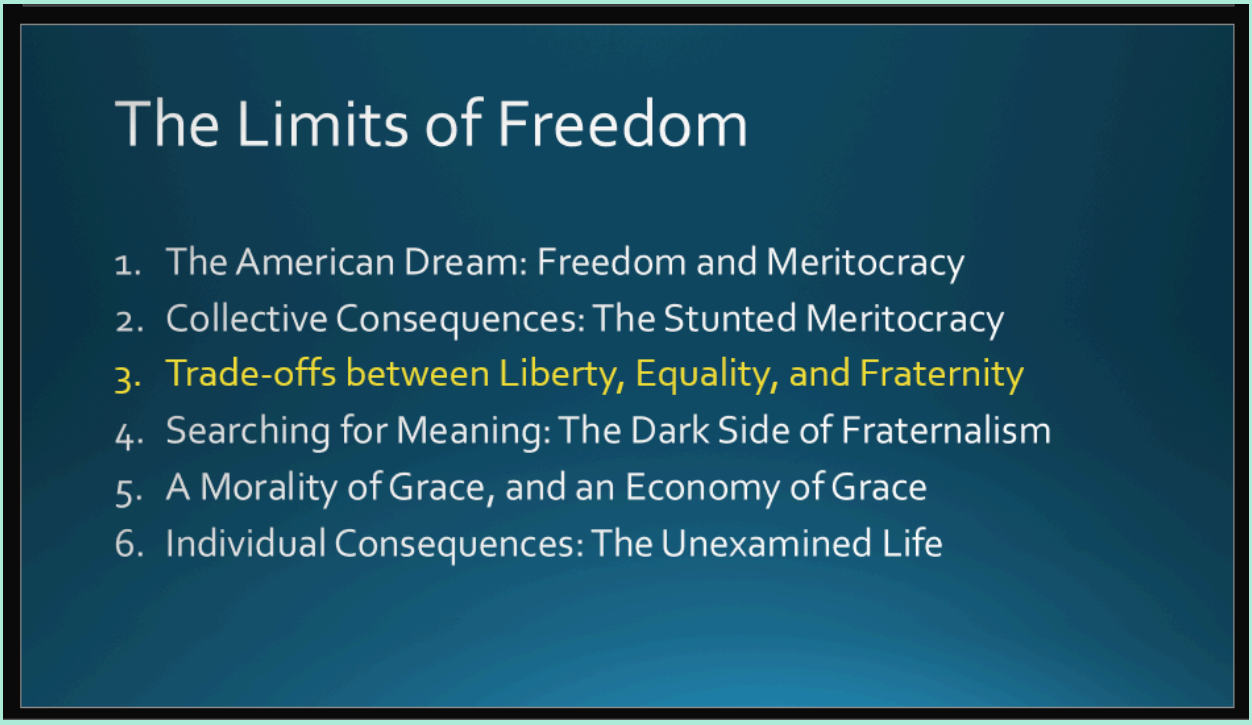
5. **Findings**: The key results of your study can be listed on one slide and then broken down across subsequent slides, with each slide covering a specific point or subpoint in your overall argument. Include slide titles that sum up each part of your argument. Collectively, the findings slides should answer the research question you posed earlier (and you may want to restate that question throughout this section). Perhaps the most important thing to remember in the findings section of a presentation—however long or short—is to *show the data*. Don't just have a list of bullet points that sum up your findings; scholars in the audience won't be convinced unless they see how you came up with those results. For quantitative studies, you will want to show relevant tables and charts that reflect the trends or patterns you discovered. Make sure the tables and charts are large enough to be easily read and digested when they appear on the screen. If you have a complex table with a slew of statistical results, for example, you should create a simplified version of that table, spread the results out across several slides, or show the full table but then break it down piece by piece across separate slides. For qualitative studies, give your audience a taste of your actual data—interesting quotes from respondents or descriptions of key episodes you ethnographically observed, captioned with some background context about each respondent or site. You will have to be judicious in picking illustrative examples, but providing this level of detail is important, as qualitative work relies heavily on this thick description to make it compelling and persuasive. If you have lengthy quotes, you generally don't want to read them word for word during your talk. You may want to put up the full quote on a slide but just focus in your comments on critical phrases or sentences, perhaps highlighting or boldfacing them on the slides.

6. **Theoretical Implications:** In scholarly presentations, you generally want to prioritize your study's *theoretical* contributions—that is, how your findings change our overall understanding of a given topic. Just like your study's findings should answer your research question, your list of theoretical implications should respond to your research problem. Say how your findings build upon or overturn past studies. If appropriate, you can also include a *policy implications* slide that goes over the practical significance of your findings—for instance, what sorts of policies or programs we might institute to address the issues raised by your study.
7. **Limitations:** This might go before the theoretical implications slide, or even be integrated into the methods slides, but at some point, you want to discuss any shortcomings of your study's approach. Was the study's data truly representative of the population of interest? If not, how might that have influenced your results? Could your data collection or analytical procedures be improved? Could we think of alternative approaches to studying this topic that could also be fruitful? Be forthright and thorough in talking through any of your study's limitations, which will inevitably prompt thinking about future research projects that you and others can take up—thereby advancing the science.
8. **Thank You *or* Questions?** Always end your presentation with a slide that thanks the audience and opens up the room to possible questions. On this slide, you should put your contact information: name, title/institution, email address, maybe even your social media information. (Remember, presentations are a great way to network with other scholars interested in the same topic.)

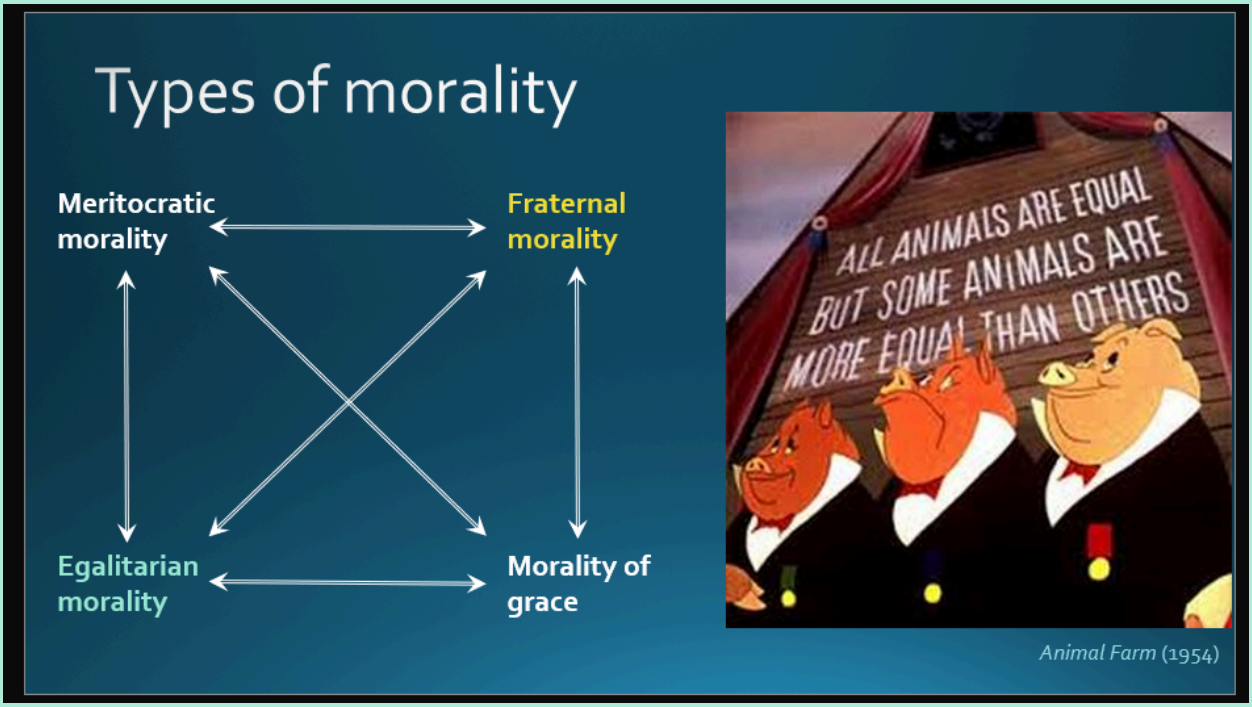


Always end your presentation with a "thank you" slide that lists your name and contact information (including social media handles, if you wish) so that audience members can reach out or stay in touch. You may also want to include slides *after* your presentation slides that provide additional information that can help you field audience questions during the Q&A, such as a demographic breakdown of your interview sample or additional charts and tables.

Deeper Dive: Tips for Presentation Slides



Especially for longer talks, it is a good idea to *signpost* where you are in the talk as you move through its various sections—perhaps by continually returning to an outline of the talk and highlighting the new section you're beginning. As can be seen in these slides from a talk Victor gave, you can do the same thing by using diagrams, highlighting different parts of the image as you move through your argument. *Halas and Batchelor* (Animal Farm)



There is a ton of advice out there (sometimes conflicting) about how to design your slides for research talks. Here, we'll mostly follow the sage guidance of the Population Reference Bureau, a nonpartisan research organization, in their concise handout, *Connecting People to Useful Information* (2003). These are the most important pointers:

1. **Follow the *rule of 6 x 6*:** You should not have more than *six bullet points* on any slide, and each bullet point should have no more than *six words*. If you have more things to say, create another slide—or, better yet, cut out the less important bullet points (you can always expand orally on what's written on the slide). For each bullet point, avoid full sentences and instead use brief phrases that will jog your memory regarding what you want to talk about at that moment—but that won't tempt the audience to spend time pouring over your slides and not paying attention to you.
2. **Keep text concise and readable:** The rule of 6 x 6 is the ideal, but frankly, most academics violate it routinely in their talks. The more important thing to remember is that you don't want to clutter your slides with walls of text, and you don't want to be reading your bullet points word for word, which will come across as robotic. To keep yourself in check, make sure the font size on your slides is not too small (say, 28 point or larger) and yet not too large (with the titles no larger than, say, 44 point).
3. **Use images to break up text:** Using images like photos, diagrams, and charts is generally a great idea. They make your visual presentation more engaging and ensure that you aren't cramming too many words on each slide. In fact, you might want to swap some of your bullet points with images that convey the gist of those points, which you then comment on in your spoken remarks. You can also have slides with images alone, which is a nice way of breaking up long sections of your talk; you can move your bullet points to the slide notes that aren't visible to the audience. You *can* overdo images in a presentation if such content doesn't fit the serious tone of a particular setting or topic, but generally speaking, more in the way of visual aids will keep your audience from dozing off. Academics generally don't have time in conference presentations to play videos, but you may find yourself including them to underscore a

particular point or to show some of your data in visual form. Note that copyright laws apply to presentation content as well, so especially if you're doing a public talk, you should try to use public domain or openly licensed images or have a good "fair use" case to incorporate that content without explicit permission. You can find relevant images on repositories dedicated to openly licensed content like Pexels and Wikimedia Commons or by searching specifically for Creative Commons-licensed material on larger photography platforms such as Flickr. (See this VCU Libraries list of sites for free-to-use images, including stock images of diverse populations.)

4. **Reiterate your research question and/or research problem across your slides:** Audience members frequently get lost in academic talks, forgetting what the research question is or why it actually matters as the presenter dives ever-deeper into their data. In oral presentations (more so than in written papers), it's a good idea to repeat yourself in order to underscore a key point that people may have missed the first time around. For instance, restate your research question as you go over your findings, and restate your research problem as you go over your theoretical implications. One useful strategy is to *signpost*—create a slide (or a slide header or footer) with an outline of the sections of your presentation, and refer back to that outline continually during your talk, highlighting or boldfacing which part of your argument you are currently covering. That outline might briefly signal what your research question or research problem is, so that the audience members always have that in mind when mulling over your data and results.
5. **Be prepared for the Q&A:** Giving the talk is only half the battle—especially when you're presenting on more controversial topics. You also have to be prepared to respond intelligently to any questions or comments that come up during the Q&A. One good strategy is to write a list of potential questions and think of possible answers beforehand. You can even place extra slides at the end of your slide deck that aren't part of your presentation proper but that you can refer to during the Q&A. This is also a good way to condense the length of your talk; for instance, if you don't want to get into the nitty-gritty of your sample's characteristics during the talk, you can skip that and tell people that you're happy to provide more details during the Q&A. You can include a table summarizing that sample information at the end of your slide deck, trotting out that slide only if people actually ask you about it.

In general, treat your slides as helpful guides to keep you on message, but not as a replacement for the talk. In fact, for most presentations, you don't want to have a prepared script of what you're going to say. Instead, learn to speak off the cuff based on concise bullet points. This takes practice, but it will pay off in reducing the labor that goes into preparing each talk and making your presentations seem more natural and conversational.

Key Takeaways

1. In research talks, you typically just have time to cover the highlights of your study, but you want to give the audience a rich sense of your data in addition to articulating the theoretical significance of your work.
2. Roundtable and poster presentations are other ways of presenting your work to fellow scholars, and they are especially good for getting feedback on preliminary work.

Exercises

1. Imagine how you might present in a poster some of the work you've done in class. What would the poster look like? What would it contain? Many helpful web resources offer advice on how to create a scholarly poster presentation. Do a web search for "scholarly poster presentation" and you'll find hundreds of sites that share tips on creating an effective poster. Visit a few of the links that your search yields. How has your vision for your poster changed, and why?
2. One way to prepare yourself for presenting your work in any format is to get comfortable talking casually with others about your research. Practice with your friends and family. Engage them in a conversation about a research project you conducted in class. If you haven't conducted research yet, talk about your research interests. Ask them to repeat what they heard you express about your proposed project. How closely do their reports match what you intended to express?

A.3. Writing Up Your Work for Other Scholars

Learning Objectives

1. Know the different parts of a journal article and what details should be included in each section.
2. Define plagiarism and explain why it should be taken seriously.



Check for updates

SOCIUS

Original Article

ASA

AMERICAN SOCIOLOGICAL ASSOCIATION

Socius: Sociological Research for a Dynamic World
Volume 5: 1–23
© The Author(s) 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/2378023119859711
srd.sagepub.com
SAGE

Industry, Firm, Job Title: The Layered Nature of Early-Career Advantage for Graduates of Elite Private Universities

Daniel Davis[1] and Amy Binder[1]

Abstract

Using concepts associated with effectively maintained inequality theory and horizontal stratification, the authors ask whether the private-public dividing line is a "threshold of consequence" for early-career market entry. To address this empirically, the authors use a novel LinkedIn data set to analyze job pathways for the graduating class of 2016 from the top 25 private and top 25 public universities in the United States. In line with past qualitative research, the authors find evidence that elite private graduates enter high-status industries in greater proportion than their public university counterparts. They also tend to get jobs at more prestigious and higher paying firms and to attain more prestigious job titles. On the basis of the evidence, the authors call for more closely analyzing the layers of advantage that may accumulate to elite graduates during key transitional moments, such as during the postgraduation job search. The authors also shed additional light on how the private-public divide is a threshold of consequence for university graduates.

Keywords

elite private and public universities, effectively maintained inequality, thresholds of consequence, college-to-career, layers of advantage

Why do wealthy families go to such lengths to land their children at a handful of elite, mostly private, universities? The question is more relevant than ever in the wake of Operation Varsity Blues, the largest admissions scandal in U.S. history, in which the Federal Bureau of Investigation charged dozens of high-wealth families in 2019 with bribing coaches and cheating on admissions exams to get their children into top schools (Medina, Benner, and Taylor 2019). One answer to the "why" question can be found in the concepts of horizontal stratification of higher education (Davies and Zarifa 2012; Gerber and Cheung 2008) and trends toward effectively maintained inequality (EMI) (Lucas 2001). Proponents of these theories envision higher education as a competitive arena of qualitatively different categories of institutions that are ranked hierarchically, from best to worst. This is not a matter of positioning colleges and universities vertically, by the level of degrees they offer, but rather points to consequential thresholds that separate more advantageous institutional segments from others on a perceived metric of quality and prestige. Independent of how large or small the actual level of difference between categories of institutions may be, universities in distinctive categories are perceived to be significantly different in status. Parents, students, and even employers use these categorical, or symbolic, boundaries to distinguish higher education institutions from one another (Alon 2009; Binder and Abel 2019; Rivera 2015b). In such an arena, affluent families scramble—legally and illegally, as it turns out—to situate their children over the threshold into a better qualitative position relative to otherwise equal competitors, with future economic and social payoffs in mind.

Horizontal stratification and EMI are, at root, explanations for how economically advantaged households seek to hoard opportunities for their sons and daughters. However, these theories also point to a related phenomenon, which is that relatively similar student contenders who might otherwise do equally well in a meritocratic competition could end up doing quite differently in the long run if they fall on one side of a consequential institutional threshold rather than the other. This echoes Merton's (1968) classic article on the Matthew effect, in which the rewards in academic science careers are rationed to a limited number. Being just below the cutoff

[1]University of California, San Diego, La Jolla, CA, USA

Corresponding Author:
Daniel Davis, University of California, San Diego, Yankelovich Center for Social Science Research, 9500 Gilman Drive, La Jolla, CA 92093, USA.
Email: dbdavis@ucsd.edu

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (http://www.creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).



point—the figurative "41st chair"—will disadvantage a scientist compared with the gains in recognition, awards, and resources enjoyed by a competitor who was only marginally better but just above the cutoff in the 40th chair. A modern-day example is featured in Bol, de Vaan, and van de Rijt's (2018) research on young scientists' early-career trajectories. Here, although reviewers' scores of the winners and losers in a grant competition were only slightly different, over the following eight years, those who had won the initial grant went on to secure more than twice as much additional funding as did those who had barely missed the threshold. Early wins produce a signaling effect to later funders, which helps these winners accumulate grants in subsequent competitions. The same could be true of graduates of private universities who cross a symbolic threshold leading to later wins.

Although rarely in conversation with theories of horizontal stratification or EMI (or, for that matter, Merton's Matthew effect), past research has indicated that graduates of private elite universities enjoy distinct advantages over their public university peers. Employers in a few premier industries use elaborate recruitment strategies only on elite campuses to hire young employees, leaving the majority of the country's college graduates on the other side of a cutoff point, without even the possibility of applying for these jobs (Binder, Davis, and Bloom 2016; Rivera 2015b). Managers at elite firms evaluate graduates from top private universities more favorably in terms of "cultural fit" with their firms' current employees than for their cognitive skills (Rivera 2012), pointing to the status signaled by elite institutional attendance. Private elite students also beat the competition in other segments of the labor market, as indicated in audit studies. Gaddis (2015) found that employers from a wide range of industries favored similar résumés from students at three elite private universities compared with a public university in the same region. In Gaddis's study, job applications containing private elite university degrees received 1.7 times more responses from employers than did evenly matched applicants from the public institution. In an older study, Ishida, Spilerman, and Su (1997) showed that the signaling effect of prestigious universities is most pronounced at the beginning of applicants' careers, when there is little experience listed on résumés apart from having attended selective colleges and universities. Paralleling the findings of Bol et al. (2018) but applied to the college sector, these studies suggest how an elite university degree can be used as a stamp of quality sought by employers (Connelly et al. 2011; Ho 2009).

Separate from studying the actions of employers, sociologists have taken students as the unit of analysis to study the effects of attendance at elite institutions. Findings are not incontrovertible, but there is suggestive evidence that elite advantage is strong in multiple social and economic realms. Elite students enter social circles that are exclusive and can lead to social closure, including in marriage and friendship networks (DiMaggio and Powell 1983). Elite private universities produce top political leaders in the United States (Zarifa and Davies 2018). Students from elite colleges and universities attend graduate school in higher numbers and in higher status departments than do students who attend public universities (Posselt and Grodsky 2017). Brand and Halaby (2006) demonstrated that elite college attendance is strongly associated not only with college graduation and graduate school attendance but also with socioeconomic status of first job and enduring socioeconomic boosts in mid- and late career. Early gaps in job taking continue to persist years later, as students graduating from private elite universities enjoy a substantial wage premium relative to their public university cohort-mates (Katchadourian and Boli 1994). Thomas and Zhang (2005) found an early-career earnings gap of 20 percent between graduates from "high-quality private" and "low-quality public" universities, while Witteveen and Attewell (2017) demonstrated that for students 4 and 10 years after graduation, a large gap persists, particularly for men. In sum, much of the sociological literature points in the direction of strong advantages for the graduates of elite private universities.

Scholarly Skepticism about Elite Advantage: Selection Effects, Sociological Blinders, and the New Economy

Although these studies point to elite advantages in the workplace, there are pockets of skepticism about whether elite private universities, in and of themselves, offer much advantage. One variant of this argument is that selection effects are the key mechanism leading to better career outcomes for students of elite colleges and universities (Dale and Krueger 2002). Although in the aggregate, private elite university attendance may be associated with higher earnings (and, by extension, better jobs), once one controls for students' high school characteristics (such as test scores), and schools to which students applied and were accepted, elite advantage in earnings washes out (Hout 2002). More recent research confirms the main findings of this argument but adds granularity: with the exception of Black, Hispanic, and lower socioeconomic status students (Dale and Krueger 2014) and women working less than full-time year round (Ge, Isaac, and Miller 2018), there are null effects of attendance at private elite universities.

A separate argument is that assumptions about elite advantage are simply overblown: although access to private elite education is plagued with inequalities, career outcomes do not reflect those same inequities down the line in the job market. An example of this argument can be found in research conducted by Brint and Yoshikawa (2017), who showed that later career workers who are employed as "C-level" executives at top companies (or who are national political leaders) are as likely to have attended public universities as private institutions, although Brint and Yoshikawa's data also showed that highly successful alumni from private and public universities cluster into different labor market sectors.

Further challenging the notion that private elite university graduates do better in the labor market is research suggesting that elite advantage is not as durable in the new economy as

In this study published in the open-access journal *Socius* (Davis and Binder 2019), you can see how the abstract is followed by the paper's (untitled) introduction, which is then followed by a literature review (titled not as "Literature Review" but more specifically based on the themes it covers). *Sage*

When you are writing an empirical study for other scholars, you generally will want to follow the format for a journal article. Reading prior literature in your area of interest is an excellent way to learn the core components of a journal article and how to write one yourself. There are also many excellent resources that can guide you as you prepare to write a journal article (see, for example, Becker 2020; Johnson et al. 2010; Moskovitz and Smith-Lovin 2017; White 2005).

Journals differ in the particular format they use, and authors have some latitude in organizing the article the way they want. That said, it's a good idea to adhere to the following sequencing and section headings so that other sociologists can easily understand your study:

1. **Abstract**
2. **Introduction** (note that published papers don't usually include this header; they just dive right into the text of the paper after the abstract).
3. **Literature Review** (journals don't typically include this header, either, but instead have sections with headings describing distinct themes or topics in the literature)
4. **Statement of the Problem** or **Problem Statement** (this may be incorporated into the end of the literature review)
5. **Hypotheses** (this also may be integrated into the end of the literature review or just implied from context)
6. **Methods** (often spread across separate sections like **Data**, **Data Collection**, **Sample**, **Case Selection**, **Measures**, and **Data Analysis**)
7. **Results** or **Findings** (journal articles often won't include this header, but will have separate sections devoted to each of the study's main findings)
8. **Discussion** and/or **Conclusion** (sometimes called **Implications** and sometimes merged into a single section called **Discussion and Conclusion**)
9. **Acknowledgments** (sometimes a footnote on the first page of an article)
10. **Declaration of Conflicts of Interest** (sometimes excluded)
11. **References** or **Works Cited**

As you can see, the sections of a journal article are roughly the same as those for research talk described earlier, but you have more space to flesh out the details, and the names of the sections will differ slightly (for instance, the slide **Theoretical Contributions** will correspond to the **Discussion** or **Conclusion** sections of a written paper). Let's review each of these sections and what should be included in them.

Nearly all empirical articles will have an *abstract*, the paragraph at the beginning of an article that summarizes the author's research question, methods used to answer the question, and key findings. Abstracts may also give you some idea of the theoretical perspective of the authors and the research problems they're attempting to address. The standard length depends on the journal, but most abstracts range between 100 and 200 words. Some journals further break down the abstract into short summaries of each section of the article. The abstract may also be followed by a list of **keywords**—terms that the article authors usually select to describe the main concepts and topics they are covering.

Again, the *introduction* does not need to be explicitly titled as such, but you will want to include at least one or two introductory paragraphs at the start of your paper. Here, you should quickly make a case for why your research is important and necessary. You might begin with general background on your topic, such as key statistics, a summary of recent trends, or historical context. You then want to quickly summarize your research problem and the research question that arises from it. (This summary may paraphrase or entirely replicate portions of your abstract, which is okay in academic papers: you want to hammer home these key points throughout the manuscript.) For instance, you could briefly describe the existing literature (saving any extended discussion for the next section), identifying where exactly your own research question and research study fit and what gaps in scholarly understanding they fill.

The size and scope of the *literature review* varies greatly depending on the discipline and the journal. Sometimes an article's literature review is published as one or more thematic sections, rather than being titled as a "literature review." Sometimes the literature review is combined with the introduction; this is less common in sociology but is quite common in applied fields like medicine, which typically have less in the way of theory to discuss. (We describe the structure and contents of the literature review in depth in **Chapter 5: Research Design**.)

The *statement of the problem* (or *problem statement*) may be a separate section of the article, or it may just be the ending paragraph or two of the literature review. In this section, you want to provide a concise description of the research

problem, the so-called *gaps* in the literature: what the current literature misses, what sort of puzzles still need to be solved, what theories could be applied to the social phenomenon of interest that would explain it better, how previous types of analysis could be extended into new areas, and so on. Here, you are taking the picture of the existing research you painted earlier in the paper and using it to make the case for the importance and necessity of your particular study.

Sometimes, researchers will follow up the literature review and problem statement with *hypotheses*. (This is more common for quantitative studies rather than qualitative ones.) Hypotheses are educated guesses as to the best answers to the research question, and they are explicitly derived from the literature review—in fact, they may be incorporated into the literature review, as shown in **Figure A.1** below. You can choose to list your hypotheses formally (e.g., numbering them and setting them off from the rest of your prose) or describe them within the main text of your manuscript.

We thus expect, all else being equal, to see an empirical relationship between state-level income inequality and class divides in parental financial investments in children:

Hypothesis 1: The gap between high- and lower-income-rank households in parental financial investments in children will be wider when state-level income inequality is higher.

MECHANISMS CONNECTING INCOME INEQUALITY AND PARENTAL INVESTMENT

Evidence of a relationship between income inequality and class gaps in parental investment would naturally raise the question of what

Figure A.1. Inserting Formal Hypotheses into Your Paper. An *American Sociological Review* article by Daniel Schneider, Orestes P. Hastings, and Joe LaBriola (2018) found that in states with higher income inequality, more affluent parents splurge on tutoring, extracurriculars, and private schools for their children. To set up the analysis, their article includes an extensive literature review that they use as the basis for a number of hypotheses. The hypotheses are formally numbered, as in the excerpt shown here, but are embedded within sections of the literature review rather than being put in a separate "Hypotheses" section (which is another option). *Sage*

The *methods* (or *methodology*) section reviews how you gathered your sample, how you measured your variables, and how you analyzed the data. You may use separate "Sample," "Case Selection," "Data," or "Data Collection" sections or subsections where you describe how you went about choosing or recruiting the people, organizations, or sites that you studied. As we mentioned in our discussion of research talks, survey-based studies, in particular, will include a discussion of specific dependent and independent variables and how they are measured. This section (sometimes titled "Measures" or something similar) will include the actual text of key survey questions, and it may also contain some

descriptive results regarding respondents' characteristics or their answers to general questions (answers that relate to the specific research question, however, will be described later in the article's "Findings" or "Results"section). In a separate "Data Analysis" section, you should describe which statistical or qualitative analytical procedures you used to process the data you collected, including the software you used and any coding strategy you adopted.

In general, you should use the methods section to describe and justify your choice of research techniques. Oftentimes, you will also want to address the limitations of your methodological approach in this section, although you may save some or all of this discussion for the end of the paper.

Table 1. Private and Public Universities, 2016 *USNWR* Rankings.

Private Universities	Public Universities
1. Princeton University	1. University of California, Berkeley
2. Harvard University	2. University of California, Los Angeles
3. Yale University	3. University of Virginia
4. Columbia University[a]	4. University of Michigan
4. Stanford University	5. University of North Carolina, Chapel Hill
4. University of Chicago	6. College of William & Mary
7. Massachusetts Institute of Technology	7. Georgia Institute of Technology
8. Duke University	8. University of California, Santa Barbara
9. University of Pennsylvania	9. University of California, Irvine
10. California Institute of Technology	9. University of California, San Diego
10. Johns Hopkins University	11. University of California, Davis
12. Dartmouth College	11. University of Illinois Urbana, Champaign
12. Northwestern University	11. University of Wisconsin, Madison
14. Brown University	14. Pennsylvania State University
15. Cornell University	14. University of Florida
15. Vanderbilt University	16. The Ohio State University
15. Washington University, St. Louis	16. University of Texas, Austin
18. Rice University	16. University of Washington
18. University of Notre Dame	19. University of Connecticut
20. Emory University	19. University of Maryland, College Park
21. Georgetown University	21. Clemson University
23. Carnegie Mellon University	21. Purdue University
23. University of Southern California	21. University of Georgia
24. Tufts University	24. University of Pittsburgh[b]
25. Wake Forest University	25. University of Minnesota, Twin Cities

Note: Rankings from 2016 *U.S. News & World Report* (*USNWR*) "Top National Universities." We separated the top 25 public universities from the top 25 private universities for our comparison. In the original *USNWR* list, public and private institutions were integrated into the same running list of hundreds of "Top National Universities."
[a]Instances of the same rank denote ties in the *USNWR* ranks. Ties are listed alphabetically.
[b]The University of Pittsburgh is technically designated a "state-related" university.

LinkedIn Hiring Data

With our list derived from *USNWR*, and partial school data derived from the College Scorecard, we then used the LinkedIn platform to collect data on the jobs that our 2016 graduates took. Among other services attracting potential users, LinkedIn provides opportunities to search for jobs, build industry connections, and display work experience and credentials. More important for our study, however, LinkedIn contains self-reported graduation and employment data for people who build profiles on the site, and the site clusters this information on what are called alumni pages for each university. Also useful for our project, the alumni page for each university offers a dashboard tool, which allows researchers to limit searchable data to only those students who reported graduating from that university in specific years. Because our interest is on an important moment of career transition—what happens when a student graduates from college and enters the labor force—we collected data for a single recent year: those who graduated from 1 of our 50 sample universities in 2016. (The data set is available upon request.) This allowed us to zero in on young users' early job experiences rather than seeing an accumulation of jobs that older users' LinkedIn profiles would contain.

This strategy is not foolproof, as even young workers may list multiple places of employment within their first year after graduation, for example, if they quickly job-hopped, worked at more than one job simultaneously, or had internships or summer jobs at these firms during college. However, this kind of amassing of job titles is substantially lower for very recent graduates than for alumni who graduated several years prior. Given this, we consider our data to be indicative of early places of employment and mostly first jobs after graduation, although not in all cases the single first job taken after attaining a baccalaureate degree. We accessed the data in the summer of 2017, allowing one full year for those who had graduated in the spring term of 2016 to assume their postcollege jobs.

There are other ways that the LinkedIn platform is an imperfect data source for scholarly research, although we have made efforts to address its limitations. Most important, we suspect that some students are likelier than others to create LinkedIn profiles, such as those in professional first jobs compared with those who are employed in less prestigious lines of

Continuing with our breakdown of this *Socius* article, we can see that the authors do not include a separate section for a statement of the research problem. Instead, they make the case for their study at the end of the literature review. Then they move into the "Methodology" section, where they first describe and justify their selection of certain universities to include in their analysis (we'll discuss case selection in **Chapter 6: Sampling**) and then begin a new subsection detailing the LinkedIn data they used. *Sage*

in past labor markets. Bessen (2015) argued that the rise of high-tech industries may contribute to a greater democratization of employment opportunities for all university graduates, as tech skills are often learned better on the job than in college classrooms. Heightened demands on public research universities to spur regional economic development have led public institutions to multiply their efforts to produce commercially viable technologies (Kenney and Mowery 2014), offer targeted and high-demand workforce training (Walshok, Munroe, and DeVries 2011), and forge new recruiting pathways with elite employers (Davis and Binder 2016). These developments could weaken the advantage of attending the most selective private universities because employers have a larger pool from which to search for employees with the highest skill sets. Hoekstra (2009) showed that attending the most selective state universities is associated with a post-graduation wage premium over less selective universities, suggesting that employers look favorably on flagship public universities. The *New York Times*, following research published by Chetty et al. (2017), recently named all of California's leading public research universities as "upward-mobility machines" (Leonhardt 2015).

Such changes in the new economy and at the nation's top public universities, in combination with Brint and Yoshikawa's (2017) findings about later executive careers held by public university graduates, could indicate that graduation from elite institutions is becoming less salient in the face of the growing technocratic demands of the current labor market. It is also true, vis-à-vis the argument about selection effects, that many of the studies that find large gaps between public and private university graduates compare the "highest quality" institutions with a wide range of less selective institutions. For example, in Gaddis's (2015) audit study, batches of graduates' résumés from the three private campuses contained 75th percentile SAT scores averaging 1540 compared with scores of 1320 for the public university campus. In such a case, selection effects cannot be ruled out.

In this study, we develop a novel data set that allows us to weigh in on the debate about private elite advantage and whether a threshold of consequence between types of universities matters for early-career outcomes. Using a well-matched set of institutions drawn from the LinkedIn platform, we show patterns of job taking among recent college graduates, which, to this point, have gone unexplored. Specifically, we show different rates of entry by the nation's top public and private university students into the nation's largest industries, firms, and job titles. Our findings support elite-advantage arguments in the literature.

Methodology

Measuring Private versus Public University Status as a Threshold of Consequence

To examine whether public or private university attendance functions as a threshold of consequence largely independent of the quality of students attending institutions, we compared job outcomes for graduating seniors in the class of 2016 at the top 25 public and top 25 private universities, as found in *U.S. News & World Report*'s (*USNWR*) ranking of "Top National Universities." (*USNWR* maintains separate lists for regional universities and liberal arts colleges, which we do not include in this analysis.) Despite misgivings about the value and accuracy of *USNWR* rankings (Espeland and Sauder 2007), we decided to use these annual reports for two reasons. First, however valid the criticisms that are lodged about the uses and abuses of *USNWR* rankings, they are the leading source of information used by the general public, prospective students, university administrators—and, potentially, employers—to learn about universities' perceived positions relative to one another. Second, we are not the only sociologists to use this ranking system. Espeland and Sauder (2007) used *USNWR* rankings in their study of law schools, in which they demonstrated the power of the rankings to radically influence professional schools' admissions policies. Likewise, Gaddis's (2015) audit study of student applications for jobs from higher and lower ranked universities and Brint and Yoshikawa's (2017) study of the educational backgrounds of later career executives and national politicians also relied in part on *USNWR* rankings.

Table 1 provides a list of the universities in our sample, while Appendix A provides information about these campuses from the National Center for Education Statistics (NCES), such as their acceptance and graduation rates, mean SAT scores, and students' average earnings 10 years after admission.[1]

Using NCES data to compare SAT scores (a measure of the "quality" of students), we find that students at the 25 top private universities and the 25 top public universities are not vastly different in ability, a least as measured by standardized tests at the 75th percentile. Because NCES data are reported in quartiles, we chose to compare scores at the 75th percentile. The 100th percentile is simply a perfect score at most institutions, which provides minimal information on differences, while the raw mean would be quite skewed because public universities admit many more students than do private universities and have much greater heterogeneity in scores. The average 75th percentile SAT score for our private university students is 1558 (789 in math, 769 in reading) compared with 1408 (728 in math, 680 in reading) for the public university students in our sample. Our comparison groups differ in performance by about 10 percent, which we find acceptable considering that both sets of scores are two full standard deviations above the national average 75th percentile, or better than 99.7 percent of the population (Staffaroni 2018). This difference also compares quite favorably with differences in test scores reported in Gaddis's (2015) audit study, in which 75th percentile scores differed by more than 200 points.

[1]Earnings data came from College Scorecard data (https://collegescorecard.ed.gov).

The *findings* or *results* section provides an in-depth discussion of what the study found—whether its hypotheses turned out to be supported by the data, and what conclusions about the phenomenon of interest can be drawn from the available evidence. For quantitative studies, you should include tables, charts, maps, and other figures that convey the relevant patterns or trends in the data; the accompanying text should break down the key findings in those tables and figures in clear prose. For qualitative studies, you should provide in-depth quotes from respondents or detailed descriptions of the spaces or events you observed. (See the chapters of the textbook devoted to quantitative and qualitative methods for more specific advice about presenting your findings in written form.)

The *discussion* section (sometimes titled "Implications" or "Discussion and Conclusion") reviews the main findings, addresses how those findings connect with the existing literature, and discusses their broader implications for research, theory, practice, or policy. Here, you should explicitly call back to the themes or authors in your literature review to make the case for how your work challenges or expands that previous scholarship. Sometimes the discussion section will end the main text of the article, and sometimes a separate *conclusion* will include another summary of the article's findings (yes, social scientists love repetition) and may even have a higher-level discussion of the study's big-picture implications.

categorized job titles into census occupational codes and work types for the Amazon and Google workplaces.

Findings

Industry: Private University Graduates Enter More Prestigious Industries at Higher Rates

The LinkedIn dashboard tool generated 74 industry categories that our sample of students entered. We reduced this number into a more manageable list of 30 industry categories, combining related industry titles. For example, our "banking and financial services" industry category is a blend of LinkedIn's "banking," "investment banking," "financial services," and "venture capital and private equity" categories. Our "retail" industry category is a combination of LinkedIn's "retail," "wholesale," and "supermarkets." Appendix E provides the full list of the 74 industries, which we reduced to 30.

To begin our analysis, we use Tableau visualizations to identify the top industries (by hiring volume) in which our 26,733 public and private university undergraduates found jobs. In this section, we focus on the top 6 industries out of our set of 30 because, as Figure 1 shows, the 24 additional industries drop off sharply in the number of students they hire directly out of our 50 universities. In a later section we briefly describe patterns in the nondominant 24 industries to see how things look similar or different in the remaining segments of the labor force.

As seen in Table 2, the same six industries show up as top employers for both private and public university graduates, suggesting that these currently are the largest employment sector "pulls" for young workers who have LinkedIn profiles coming out of top universities in the United States. Graduation from a private versus a public university is statistically significantly associated with differing rates of entrance into these dominant industries, as evidenced by a χ^2 test of independence ($\chi^2[5, N = 19{,}125] = 1{,}185.73$, $p < .001$). Private university graduates are overrepresented in these top sectors, especially in the most prestigious of them.

The three most entered industries by private elite university students are (1) computers and Internet, (2) management consulting, and (3) banking and financial services. This matches qualitative studies revealing that elite students perceive high tech, banking, and management consulting as the most desirable and prestigious job destinations in today's economy (Binder et al. 2016; Roose 2014). The next leading industry choices for workers from private universities are (4) healthcare, (5) accounting, and (6) aerospace.

For the public university graduates in our sample, (1) computers and Internet also take top billing as the industry hiring the largest number of students, and (2) the hospital and healthcare industry hires the second largest volume of recent graduates. Although the "hospital and healthcare" sector is included in the top six industries for both private and public job seekers, it ranks second for public graduates but sixth for private graduates; this is potentially also because "hospital and healthcare" does not include medical school, which LinkedIn categorizes under graduate school.[3]

Next comes (3) "accounting," which consists mainly of jobs with the "big four" large auditing firms: Deloitte, KPMG, Ernst & Young, and PricewaterhouseCoopers. For the public university graduates, management consulting shows up fourth on the list, but it is noticeably dominated by jobs at Deloitte, which claimed 63 percent of the public university workers in the "management consulting" industry overall (compared with only 31 percent of the private university students in management consulting). Although Deloitte does do consulting work, and is listed by LinkedIn in the "management consulting" category, it more frequently is categorized as an auditing and accounting firm; in fact, Deloitte is the largest of the "big four" accounting firms (http://big-4accountingfirms.org). It is, in any case, along with other firms in the "accounting" industry, a top employer for public university graduating seniors but less so for private university graduates. The next industries hiring the largest numbers of top public university students are (5) airlines and aerospace and, finally, (6) banking and financial services, an industry that, as we saw previously, is dominated by graduating seniors from private universities.

Overall, the early picture that emerges from examining LinkedIn data for industry entry is that private university and public university students who have LinkedIn profiles enter these six leading industries at different rates, and that private university students exhibit a strong advantage in certain industries (particularly banking and consulting) when one takes into account the much smaller number of graduating seniors from private institutions. Because these are high-paying industries (Roose 2014), this is partial evidence that private university advantage may be an enduring fact in the today's labor market. Yet as we demonstrate in the next two sections, private university graduates also experience advantages by entering more highly ranked firms within several of the top six industries compared with their public university peers, and they occupy more high-profile job titles, even in the technology field, which at first glance (see Table 1) appears to be a more equitable playing field.

Firm: Private University Graduates Enter More Highly Ranked Firms within Each Industry

Management Consulting versus Accounting. Looking more closely at specific firms within our 30 industries provides a

[3]Because graduate school is different from the regular labor market, we excluded it from our list of top seven. However, it is useful to note that graduate school (which includes professional schools) was the third most frequent destination directly out of school for private elite students, but only the eighth most frequent destination for public university graduating seniors.

After its “Methodology” section, the *Socius* article organizes its findings into specific subsections, each focused on a *specific* argument the authors are making about their data—note the use of nouns and verbs in the headers. The paper backs up its arguments with numerous tables and charts (not shown). Rather than having a separate “Discussion” section, the authors chose to end their article with a single “Conclusion” section covering all the paper’s higher- and lower-level implications and discussing directions for future research. *Sage*

staffing and human resources workers needed to fill these high-turnover positions—all of which are more commonly held jobs among public university graduates (see Table 7).

Patterns in Less Dominant Industries: Overrepresentation of Public University Graduates

Having demonstrated what patterns look like in four of the six dominant industries hiring from our sample (banking and financial services, management consulting, accounting, and computer and Internet), we can also use our data set to comment on noteworthy trends in the nondominant industries in the LinkedIn sample, that is, the 24 industries that make up the rest of our sample. Although we do not go through these industries with the same precision as in earlier sections, we note that all findings reported here are statistically significant.

In the telecommunications and wireless industry, firms such as AT&T and Qualcomm hired exclusively public university students from our 2016 sample. Biotechnology and medical device firms such as Eli Lilly, Illumina, and Genentech hired 96 percent of the sample from public universities. Pharmaceutical companies such as Bayer and Pfizer hired 90 percent public university students, while manufacturing and construction firms such as John Deere and Turner Construction hired 85 percent of their 2016 employees from public universities. Automotive firms such as Ford and GM hired 84 percent of new employees from our sample from public universities. Many of these are excellent jobs in good industries, offering professional career ladders. Yet we find that these nondominant industries reveal within-industry hierarchies that speak to some of our earlier findings, though to a lesser extent than in the dominant industries because they hire fewer workers. Although we do not have the space in this article to look at these hierarchies in greater detail, for example, in the automotive industry, whereas Ford and GM employed a small percentage of workers from private universities, the more prestigious Tesla hired workers only from private universities. The less prestigious Michelin tire company and Enterprise Rent-A-Car employed workers only from public universities.

A few nondominant industries from the sample were more likely to hire graduates from private universities. Again, although we do not go into great depth here, some of these trends bolster our earlier findings about private elite graduates' gravitation toward seemingly elite segments. In the international affairs and development industry, for example, agencies such as the U.S. Department of State, the World Bank, and the International Monetary Fund hired 89 percent of their new employees from private universities. The remaining 11 percent of workers in this industry came from public universities; all of the public graduates went into the Peace Corps. Education management shows similar patterns. Public universities sent 1.4 percent of their graduates into this industry, while private universities sent a similarly small 2.3 percent of their graduates into the sector. However, 84 percent of those hired by Teach for America (TfA) in this sector came from private universities. Although not "education management" in the traditional sense, we placed TfA in this category rather than with permanent teachers for multiple reasons: no teaching credential is required, there are alternative pay and placement structures, and there are atypical continuation rates as teachers after completion of the TfA program. In more traditional education sectors, teaching positions in grade schools and high schools were filled largely by graduates of public universities, a robust 88 percent. Perhaps not surprisingly, students from top private universities who choose to go into education rarely enter career pathways to become permanent teachers; they either sign up with TfA—and more than 80 percent leave their positions within three years (Kavanagh and Dunn 2013)—or take other jobs in a separate education management sector.

Public university graduates in our sample overwhelmingly filled the retail-oriented industry jobs reported in the LinkedIn data set. Of the 774 employees working for companies such as Walgreen's, Starbucks, and Home Depot, 97 percent graduated from public universities. Of the 185 employees in consumer goods, such as Coca-Cola and Black & Decker, 81 percent were from public universities. Government administration (local, state, and federal) is more equitable, drawing 59 percent of recent graduates from public universities and 41 percent from private universities. However, of these numbers, the Centers for Disease Control and Prevention (CDC) claimed more than half of all of the private university early career workers in government administration, and all graduated from Emory, which physically neighbors the center. Removing the CDC from the sample, we found that 73 percent of those in the government administration industry graduated from public universities; these workers took jobs in city and county governments, the U.S. Food and Drug Administration, the U.S. Census Bureau, state legislatures, and staff positions in the U.S. House of Representatives. Notably, only students from private universities in our sample—11 employees—obtained positions working in the U.S. Senate, which is arguably a more prestigious governmental destination.

Conclusion

Our examination of LinkedIn data, although an imperfect data source, suggests that students from top private universities and top public universities have different pathways into the labor market in terms of the industries and firms they enter and even their job title placements within the same firms. Notably, the highest paying of the top industries—banking and financial services, management consulting, and computers and Internet—have an overrepresentation of private elite graduates. Our findings reinforce Gaddis's (2015) audit study of a wide range of employers, as well as qualitative work studying elite employers (Binder et al. 2016; Ho 2009; Rivera 2015a), which shows that top private graduates

Somewhere in the discussion or conclusion there will also be a discussion of **limitations**, which describes the weaknesses of the study's methodological or theoretical approach. As we emphasized earlier, good researchers are always candid about how and where their research may fall short, and all sociological research has limitations, given the complexity of what we study. In this section, the authors may explicitly discuss the **scope conditions** of the research—under what conditions the research findings hold, and under what conditions they do not.

Immediately following the main text of the article there may be a *declaration of conflicts of interest* section and an *acknowledgments* section. As we discussed in **Chapter 8: Ethics**, scholars need to be upfront about any financial or other personal benefits they receive that relate to the topics they study, and including a separate declaration section encourages this transparency. In the acknowledgments section at the end of the paper (which is sometimes included as a footnote at the top of the article instead), you should personally thank anyone who read through drafts of your work or provided relevant advice. You should also thank any formal peer reviewers, even if they were anonymous ("We thank the three anonymous reviewers for their helpful comments"). Importantly, the acknowledgments section should also indicate the sources of any funding that were used to complete the study, such as the names of grants and the institutions that provided them.

At the end of the article will be a list of *references* (also titled "Works Cited"). Reports written for scholarly consumption also include in-text citations throughout the report that correspond to those references (as we describe further in **Table A.1**). Before or after the **reference list**, you may also include an *appendix* (or multiple appendices) with supplementary tables, figures, or other information providing additional context or evidence of the rigor of the study's analysis.

Appendix B

Data-Gathering Strategy for the Amazon and Google Case Studies

To view the individual profiles of LinkedIn users working at Amazon and Google, we spent several weeks requesting more than 3,000 users who work at these two firms to accept our LinkedIn "connection." LinkedIn allows users to see the profiles of not only their direct connections but also the connections of connections up to two degrees of separation. More than 1,000 Amazon and Google workers from our case campuses accepted our connection, allowing us to see virtually every Amazon and Google worker's profile from our 50 case campuses who graduated in 2016: 1,343 graduates in all.

We examined the pages of only those who had graduated with their bachelor's degrees in 2016 and subsequently took jobs at Amazon or Google in the calendar year 2016 or 2017. We discarded the rest. We gathered job titles and then matched those titles with census occupational codes. There were a total of 253 distinct job titles in the LinkedIn pages, which we matched to 56 census occupational codes (see Appendix D for a breakdown). We then further collapsed those 56 codes into 11 distinct types of work (see Appendix E for a breakdown). This allowed us to have a standardized way to compare the types of work done at these two companies by graduates of private versus public universities.

Appendix C

Amazon and Google Job Titles, by Census Occupational Codes and Titles

20. General and Operations Managers
- Area manager
- Business manager
- Program manager
- Shift manager
- Supply chain manager

40. Advertising and Promotions Managers
- Marketing manager
- Media manager

50. Marketing and Sales Managers
- Associate product marketing manager
- Content marketing manager
- Digital marketing manager
- Digital strategy lead
- Product manager
- Project marketing manager
- Strategic account manager
- Strategic partner manager

60. Public Relations and Fundraising Managers
- Community manager

110. Computer and Information Systems Managers
- Associate team lead
- GIS operations lead
- IT team leader
- Partner technology manager

140. Industrial Production Managers
- Group manager
- Junior interstock manager
- Manager I
- Manager II
- Quality area manager
- Senior product manager

150. Purchasing Managers
- Associate vendor manager
- Authorization for expenditure manager
- Vendor manager

160. Transportation, Storage, and Distribution Managers
- Delivery station liaison
- Inbound area manager

310. Food Service Managers
- Instock manager at Amazon Fresh

520. Wholesale and retail buyers, except farm products
- Junior sales coordinator
- Strategic vendor specialist

540. Claims Adjusters, Appraisers, Examiners, and Investigators
- Compensation operations coordinator

565. Compliance Officers
- Interim quality assurance advisor
- Payments compliance

630. Human Resource Workers
- Administrative business partner
- Associate lead
- Associate staffing channels specialist
- Associate, executive recruiting
- Connections advisor
- Diversity specialist
- Engagement coordinator
- Engineering sourcer
- Hardware staffing program manager and strategy specialist
- Hiring event coordinator
- HR business partner
- Lead recruiting coordinator
- Machine learning recruiting specialist
- Online candidate specialist
- People analyst
- People operations coordinator

The *Socius* article contains multiple appendices at the end that provide supplementary details about the data and the authors' data collection strategy. The final pages of the article include an acknowledgments section that mentions the study's source of funding and thanks a number of people, including reviewers. It ends with the ASA-formatted list of references. (Note that "ORCID iD" refers to a classification system for identifying specific scholars.) *Sage*

Acknowledgments

We would like to thank the Kauffman Foundation for generous funding. We would also like to thank April Sutton, Edward Hunter, Kevin Beck, and our reviewer and editors for valuable comments on earlier drafts. Additional thanks to Carey Blakely for editing support and to Marialexia Zaragoza, Ian Barrett, and Rachel Davis for data collection support.

ORCID iD

Daniel Davis https://orcid.org/0000-0001-6915-0523

References

Alon, Sigal. 2009. "The Evolution of Class Inequality in Higher Education: Competition, Exclusion, and Adaptation." *American Sociological Review* 74(5):731–55.

Avadhut. 2019. "The Big Four Accounting and Consulting Firms: Jobs, Salaries, and Internships." Retrieved June 14, 2019. https://www.financewalk.com/big-four-accounting-firms-salaries-jobs-internships/.

Bain & Company. 2019. "Bain at University of Virginia (UVA)." Retrieved June 14, 2019. https://careers.bain.com/bainonyourcampus/SchoolDetails?3_115_3=13936&id=213&contactId=213.

Bessen, James. 2015. *Learning by Doing: The Real Connection between Innovation, Wages, and Wealth*. New Haven, CT: Yale University Press.

Binder, Amy, Daniel Davis, and Nick Bloom. 2016. "Career Funneling: How Elite Students Learn to Define and Desire 'Prestigious' Jobs." *Sociology of Education* 89(1):20–39.

Binder, Amy, and Andrea Abel. 2019. "Symbolically Maintained Inequality: How Harvard and Stanford Students Construct Boundaries among Elite Universities." *Sociology of Education* 92(1):41–58.

Bol, Thijs, Mathijs de Vaan, and Arnout van de Rijt. 2018. "The Matthew Effect in Science Funding." *Proceedings of the National Academy of Sciences of the United States of America* 115(19):4887–90.

Brand, Jennie E., and Charles N. Halaby. 2006. "Regression and Matching Estimates of the Effects of Elite College Attendance on Educational and Career Achievement." *Social Science Research* 35(3):749–70.

Brint, Steven, and Sarah R. K. Yoshikawa. 2017. "The Educational Backgrounds of American Business and Government Leaders: Inter-industry Variation in Recruitment from Elite Colleges and Graduate Programs." *Social Forces* 96(2):561–90.

Brown, Mike. 2018. "The Drawbacks and Deceptions of LinkedIn." Retrieved November 28, 2018. https://lendedu.com/blog/drawbacks-deceptions-linkedin/.

Butcher, Dan. 2018. "Here's How Much You'll Make at McKinsey, Bain and BCG in the U.S." Retrieved June 14, 2019. https://news.efinancialcareers.com/us-en/246284/work-in-management-consulting-how-much-you-will-make-at-mckinsey-bcg-and-bain.

Clark, Dorie. 2014. "Where College Students Want to Work—And Why." Retrieved December 28, 2018. https://www.forbes.com/sites/dorieclark/2014/04/02/where-college-students-want-to-work-and-why/68ee5b4b7cf4.

Chetty, Raj., John N. Friedman, Emmanuel Saez, Nicholas Turner, and Danny Yagan. 2017. "Mobility Report Cards: The Role of Colleges in Intergenerational Mobility." *NBER* Working Paper No. 23618. Retrieved September 16, 2018. http://www.equality-of-opportunity.org/papers/coll_mrc_paper.pdf.

Connelly, Brian L., S. Trevis Certo, R. Duane Ireland, and Christopher R. Reutzel. 2011. "Signaling Theory: A Review and Assessment." *Journal of Human Resources* 49(2):323–58.

Dale, Stacy Berg, and Alan B. Krueger. 2002. "Estimating the Payoff to Attending a More Selective College: An Application of Selection on Observables and Unobservables." *Journal of Economics* 117(4):1491–1527.

Dale, Stacy Berg, and Alan B. Krueger. 2014. "Estimating the Effects of College Characteristics over the Career Using Administrative Earnings Data." *Journal of Economics* 117(4):1491–1527.

Davies, Scott, and David Zarifa. 2012. "The Stratification of Universities: Structural Inequality in Canada and the United States." *Research in Social Stratification and Mobility* 30(2):143–58.

Davis, Daniel, and Amy Binder. 2016. "Selling Students: The Rise of Corporate Partnership Programs in University Career Centers." *Research in the Sociology of Organizations* 46:395–422.

DiMaggio, Paul J., and Walter W. Powell. 1983. "The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields." *American Sociological Review* 48(2):147–60.

Espeland, Wendy N., and Michael Sauder. 2007. "Rankings and Reactivity: How Public Measures Recreate Social Worlds." *American Journal of Sociology* 113(1):1–40.

Friedman, Thomas. 2014. "How to Get a Job at Google." Retrieved June 14, 2019. https://www.nytimes.com/2014/02/23/opinion/sunday/friedman-how-to-get-a-job-at-google.html.

Gaddis, Michael S. 2015. "Discrimination in the Credential Society: An Audit Study of Race and College Selectivity in the Labor Market." *Social Forces* 93(4):1451–79.

Ge, Suqin, Elliott Isaac, and Amalia Miller. 2018. "Elite Schools and Opting-In: Effects of College Selectivity on Career and Family Outcomes." NBER Working Paper No. 25315. Retrieved December 26, 2018. https://www.nber.org/papers/w25315.pdf.

Gerber, Theodore P., and Sin Yi Cheung. 2008. "Horizontal Stratification in Postsecondary Education: Forms, Explanations, and Implications." *Annual Review of Sociology* 34:299–318.

Guillory, Jamie, and Jeffrey T. Hancock. 2012. "The Effect of LinkedIn on Deception in Resumes." *Cyberpsychology, Behavior, and Social Networking* 15(3):135–40.

Ho, Karen. 2009. *Liquidated: An Ethnography of Wall Street*. Durham, NC: Duke University Press.

Hoekstra, Mark. 2009. "The Effect of Attending a Flagship State University on Earnings: A Discontinuity-Based Approach." *Review of Economics and Statistics* 91(4):717–24.

Hout, Michael. 2002. "Social and Economic Returns to College Education in the United States." *Annual Review of Sociology* 38:379–400.

Hout, Michael, Tom W. Smith, and Peter V. Marsden. 2014. "Prestige and Socioeconomic Scores for the 2010 Census Codes: GSS Methodological Report No. 124." Retrieved December 29, 2018. http://gss.norc.org/Documents/reports/methodological-reports/MR124.pdf.

IGotAnOffer. 2018. "The 9 Most Prestigious Consulting Firms in the World." Retrieved June 14, 2019. https://igotanoffer.com/blogs/mckinsey-case-interview-blog/top-consulting-firms.

The sections we've just described are relevant to *empirical* journal articles. As we mentioned in **Chapter 5: Research**

Design, other types of articles include theoretical and review articles, which tend to be organized into sections based on the themes, topics, or arguments they cover. Scholarly books have a looser structure as well, though those meant for academic audiences may roughly follow the sequence of a journal article—with one chapter devoted to the introduction, another to the literature review, another to the study's methods, and so on.

If you already have a journal in mind that you want to submit to, it's a good idea to look up their *submission guidelines* (just do a web search for the target journal and look for the page on its site regarding submissions). The journal's submissions guidelines should go over the format you need to use in writing your manuscript, including any requirements in terms of maximum word count for your manuscript, the word limit for the abstract, the number of keywords to provide, and the citation style you should use for your references.

Citing Your Sources



South and Lei 11

unique determinants, it is possible that some hard-to-quantify change in the young adult cultural zeitgeist is driving changes in these proximate determinants of casual sexual activity, as well as trends in casual sex. Growing individualism and reduced sociability might lead to less partying (and hence less drinking), more computer gaming, and less autonomous living, while also diminishing the desire for sexual intercourse—at least the type of casual encounters captured in this analysis. Causation could also run in the reverse direction if a diminished desire for casual sex leads youth to party, and drink, less frequently and to play more computer games, perhaps all the while living in the proverbial parents' basement. Quantifying these potentially distal sources of change in young adults' casual sexual activity is likely to be difficult, so qualitative studies may have much to offer here.

Future research might also profit by redressing some of the limitations of this analysis. The small sample size makes it difficult to detect significant associations or subgroup differences. Our measure of casual sexual activity is rather crude and could both undercount and overcount the nonromantic sexual encounters considered to be casual sex. The measure is also insensitive to heterogeneity in the types of these encounters. Trends in, and determinants of, one-time sex with strangers might differ substantially from sexual encounters between friends or former romantic partners and from encounters that one or both participants hope will lead to a more serious romantic relationship. College "hookups" might be a unique subtype of casual sexual encounters driven by a distinct set of factors (Allison 2016; England and Ronen 2015).

We note as well that our analysis leaves unexplained a substantial portion of the decline in young adults' casual sexual behavior, particularly among young women. Trends in the hypothesized mediating factors included in this analysis explain more than half of the decline in young men's odds of engaging in casual sex but account for only about one quarter of the decline in young women's probability of having a nonromantic sexual encounter. Further research is needed to identify additional causes of the decline in casual sexual activity among young adults. Perhaps the intensifying concern with interpersonal sexual violence and sexual coercion as exemplified in the #MeToo movement has begun inhibiting presumably voluntary casual sexual encounters between young women and men. The impact of this and other broad cultural shifts will also likely be difficult to measure but may well require consideration in order to develop a comprehensive assessment of the decline in young adults' casual sexual activity.

ORCID iDs

Scott J. South https://orcid.org/0000-0002-0006-5133
Lei Lei https://orcid.org/0000-0003-2161-1217

References

Abma, Joyce C., and Gladys M. Martinez. 2017. "Sexual Activity and Contraceptive Use among Teenagers in the United States, 2011–2015." National Health Statistics Reports No. 104. Retrieved February 11, 2021. https://www.cdc.gov/nchs/data/nhsr/nhsr104.pdf.

Allison, Rachel. 2016. "Family Influences on Hooking Up and Dating among Emerging Adults." *Sexuality & Culture* 20(3): 446–63.

Allison, Rachel, and Barbara J. Risman. 2017. "Marriage Delay, Time to Play? Marital Horizons and Hooking Up in College." *Sociological Inquiry* 87(3):472–500.

Armstrong, Elizabeth A., Paula England, and Alison C. K. Fogarty. 2012. "Accounting for Women's Orgasm and Sexual Enjoyment in College Hookups and Relationships." *American Sociological Review* 77(3):435–62.

Arnett, Jeffrey Jensen. 2004. *Emerging Adulthood: The Winding Road from the Late Teens through the Twenties*. New York: Oxford University Press.

Arnett, Jeffrey Jensen. 2018. "Getting Better All the Time: Trends in Risk Behavior among American Adolescents since 1990." *Archives of Scientific Psychology* 6:87–95.

Auster, Carol J., Caroline L. Faulkner, and Hannah J. Klingensein. 2018. "Hooking Up and the 'Ritual Retelling': Gender Beliefs in Post-hookup Conversations with Same-Sex and Cross-Sex Friends." *Socius* 4. Retrieved February 11, 2021. https://journals.sagepub.com/doi/10.1177/2378023118807044.

Bauman, Eric P., and Scott J. South. 2001. "Community Effects on Youth Sexual Activity." *Journal of Marriage and Family* 63(2):540–54.

Bersamin, Melina M., Byron L. Zamboanga, Seth J. Schwartz, M. Brent Donnellan, Monika Hudson, Robert S. Weisskirch, Su Yeong Kim, et al. 2014. "Risky Business: Is There an Association between Casual Sex and Mental Health among Emerging Adults?" *Journal of Sex Research* 51(1):43–51.

Bogle, Kathleen A. 2008. *Hooking Up: Sex, Dating, and Relationships on Campus*. New York: NYU Press.

Breen, Richard, Kristian Bernt Karlson, and Anders Holm. 2013. "Total, Direct, and Indirect Effects in Logit and Probit Models." *Sociological Methods & Research* 42(2):164–91.

Curtis, Tyrone J., Nigel Field, Soazig Clifton, and Catherine H. Mercer. 2018. "Household Structure and Its Association with Sexual Risk Behaviours and Sexual Health Outcomes: Evidence from a British Probability Sample Survey." *BMJ Open* 8(12):e024255.

England, Paula, and Shelly Ronen. 2015. "Hooking Up and Casual Sex." Pp. 192–96 in *International Encyclopedia of the Social and Behavioral Sciences*. 2nd edition. New York: Elsevier.

Ethier, Kathleen A., Laura Kann, and Timothy McManus. 2018. "Sexual Intercourse among High School Students—29 States and United States Overall, 2005–2015." *Morbidity and Mortality Weekly Report* 66(51 & 52):1393–97.

Garcia, Justin R., Chris Reiber, Sean G. Massey, and Ann M. Merriwether. 2012. "Sexual Hookup Culture: A Review." *Review of General Psychology* 16(2):161–76.

Grant, Gary, Elaine M. Eshbaugh, and Jacquelyn Wiersma. 2008. "Sex for You, but Not for Me: Discontinuity in Undergraduate

This *Socius* article, "Why Are Fewer Young Adults Having Casual Sex" (South and Lei 2021), has a nicely formatted—if very *un*sexy—list of references at the end, written in American Sociological Association (ASA) style. Note that unlike American Psychological Association (APA) style, which is the most popular citations style in the social sciences, ASA style includes first names in its reference lists and uses "and" instead of ampersands (amen). *Sage*



12 Socius: Sociological Research for a Dynamic World

Emerging Adults' Definitions of 'Having Sex.'" *Journal of Sex Research* 45(4):329–37.

Horowitz, Ava D., and Louise Spicer. 2013. "'Having Sex' as a Graded and Hierarchical Construct: A Comparison of Sexual Definitions among Heterosexual and Lesbian Emerging Adults in the U.K." *Journal of Sex Research* 50(2):139–50.

Ingraham, Christopher. 2019. "The Share of Americans Not Having Sex Has Reached a Record High: You Can Blame Young People for the Dry Spell, Data Show." *The Washington Post*, March 29.

Jabr, Ferris. 2019. "Can You Really Be Addicted to Video Games?" *The New York Times Magazine*, October 27.

Johnson, Matthew D. 2013. "Parent-Child Relationship Quality Directly and Indirectly Influences Hooking Up Behavior Reported in Young Adulthood through Alcohol Use in Adolescence." *Archives of Sexual Behavior* 42(8):1463–72.

Julian, Kate. 2018. "The Sex Recession." *The Atlantic* 322(5): 78–94.

Karlson, Kristian Bernt, Anders Holm, and Richard Breen. 2012. "Comparing Regression Coefficients between Same-Sample Nested Models Using Logit and Probit: A New Method." *Sociological Methodology* 42(1):286–313.

Katz, Lawrence F., and Alan B. Krueger. 2019. "Understanding Trends in Alternative Work Arrangements in the United States." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 5(5):132–46.

Kohler, Ulrich, Kristian Bernt Karlson, and Anders Holm. 2011. "Comparing Coefficients of Nested Nonlinear Probability Models." *The Stata Journal* 11(3):420–38.

Lei, Lei, and Scott J. South. 2021. "Explaining the Decline in Young Adult Sexual Activity in the United States." *Journal of Marriage and Family* 83(1):280–95.

Lopresi, Pamela, Shayne Spaulding, and Demetra Smith Nightingale. 2019. "Disconnected Young Adults: Increasing Engagement and Opportunity." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 5(5):221–43.

Lyons, Heidi A., Wendy Manning, Peggy Giordano, and Monica Longmore. 2013. "Predictors of Heterosexual Casual Sex among Young Adults." *Archives of Sexual Behavior* 42(4):585–93.

Lyons, Heidi A., Wendy D. Manning, Monica A. Longmore, and Peggy Giordano. 2014. "Young Adult Casual Sexual Behavior: Life-Course-Specific Motivations and Consequences." *Sociological Perspectives* 57(1):79–101.

Manning, Wendy D., Monica A. Longmore, and Peggy C. Giordano. 2005. "Adolescents' Involvement in Non-romantic Sexual Activity." *Social Science Research* 34(2):384–407.

Manning, Wendy D., Peggy C. Giordano, and Monica A. Longmore. 2006. "Hooking Up: The Relationship Contexts of 'Nonrelationship' Sex." *Journal of Adolescent Research* 21(5):459–83.

Owen, Jesse, Frank D. Fincham, and Jon Moore. 2011. "Short-Term Prospective Study of Hooking Up among College Students." *Archives of Sexual Behavior* 40(2):331–41.

NIDA (National Institute on Drug Abuse). 2018. "Monitoring the Future Survey: Trends in Prevalence of Various Drugs." Retrieved from https://www.drugabuse.gov/publications/drug-facts/monitoring-future-survey-high-school-youth-trends on 2019, December 1.

PSID (Panel Study of Income Dynamics). 2016. *PSID Transition into Adulthood Supplement 2015 User's Guide*. Ann Arbor: Institute for Social Research, University of Michigan.

Payne, Krista K. 2019. "Young Adults in the Parental Home, 2007 & 2018." Family Profiles, FP-19-04. Bowling Green, OH: National Center for Family & Marriage Research.

Rubin, Donald B. 1987. *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley.

Sassler, Sharon, Desiree Ciambrone, and Gaelan Benway. 2008. "Are They Really Mama's Boys/Daddy's Girls? The Negotiation of Adulthood upon Returning to the Parental Home." *Sociological Forum* 23(4):670–98.

South, Scott J., Dana L. Haynie, and Sunita Bose. 2005. "Residential Mobility and the Onset of Adolescent Sexual Activity." *Journal of Marriage and Family* 67(2):499–514.

Tillman, Kathryn Harker, Karin L. Brewster, and Giuseppina Valle Holway. 2019. "Sexual and Romantic Relationships in Young Adulthood." *Annual Review of Sociology* 45:133–53.

Townsend, John Marshall, Peter K. Jonason, and Timothy H. Wasserman. 2020. "Associations between Motives for Casual Sex, Depression, Self-Esteem, and Sexual Victimization." *Archives of Sexual Behavior* 49(4):1189–97.

Twenge, Jean M., Ryne A. Sherman, and Brooke E. Wells. 2017a. "Declines in Sexual Frequency among American Adults, 1989–2014." *Archives of Sexual Behavior* 46(8):2389–2401.

Twenge, Jean M., Ryne A. Sherman, and Brooke E. Wells. 2017b. "Sexual Inactivity during Young Adulthood Is More Common among U.S. Millennials and iGen: Age, Period, and Cohort Effects on Having No Sexual Partners after Age 18." *Archives of Sexual Behavior* 46(2):433–40.

Ueda, Peter, Catherine H. Mercer, Cyrus Ghaznavi, and Debby Herbenick. 2020. "Trends in Frequency of Sexual Activity and Number of Sexual Partners among Adults Aged 18 to 44 Years in the US, 2000–2018." *JAMA Network Open* 3(6): e203833.

Vander Ven, Thomas, and Jeffrey Beck. 2009. "Getting Drunk and Hooking Up: An Exploratory Study of the Relationship between Alcohol Intoxication and Casual Coupling in a University Sample." *Sociological Spectrum* 29(5):626–48.

Weitzman, Abigail. 2020. "The Social Production and Salience of Young Women's Desire for Sex." *Social Forces* 98(3): 1370–1401.

Wilcox, W. Bradley, and Samuel Sturgeon. 2018. "Too Much Netflix, Not Enough Chill: Why Young Americans Are Having Less Sex." *Politico*, February 8. Retrieved November 20, 2019. https://www.politico.com/magazine/story/2018/02/08/why-young-americans-having-less-sex-216953/.

Author Biographies

Scott J. South is Distinguished Professor of Sociology at the University at Albany, State University of New York. His recent research examines racial and ethnic differences in neighborhood attainment, relationship formation among adolescents and young adults, and the social consequences of imbalanced sex ratios.

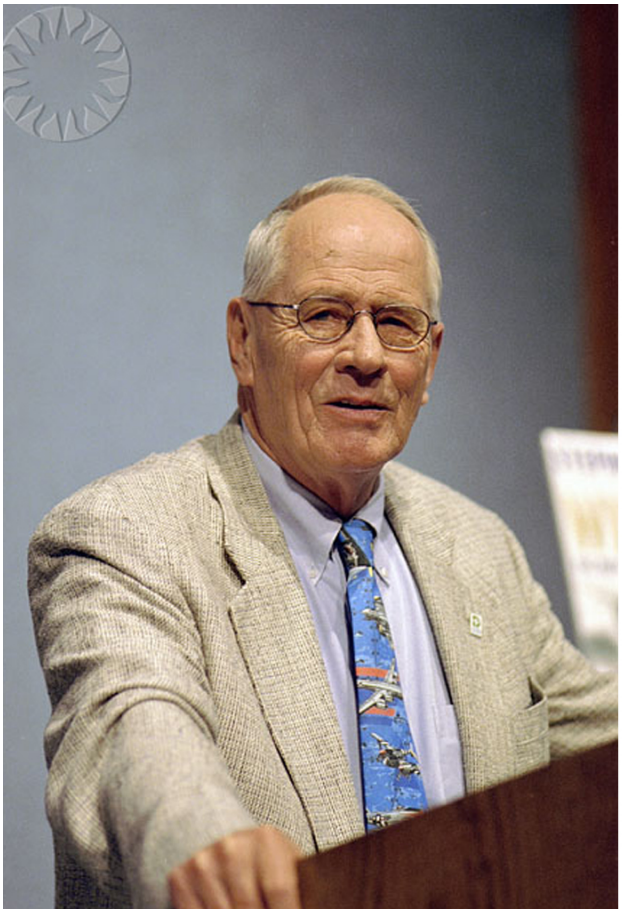
Lei Lei is an assistant professor in the Department of Sociology at Rutgers University. Her research interests include social determinants of health, family dynamics, and social inequality. She has published articles examining neighborhood effects on children's health and educational outcomes, demographic behaviors during young adulthood, and the health of migrants and left-behind family members.

Researchers use citations to disclose the sources of their facts and indicate that evidence from other people's work backs up the specific claims they are making. They also use citations when they mention ideas that aren't their own. It is critical to acknowledge whenever you are borrowing other people's ideas by properly and fully citing those authors in your paper. Otherwise, you may be guilty—even unwittingly—of plagiarism (i.e., the presentation of someone else's work as your own).

Note that just including the source in your list of references is not sufficient. You must include an in-text citation wherever you bring up the other person's idea. And even if you cite the relevant authors, you can still be guilty of plagiarizing them if you used too much of their wording. (This is called **mosaic plagiarism**, and it can also get you in big trouble.) When you paraphrase someone's writing, you need to do more than just change some words and reorder the phrases; you need to convey the ideas in your own words, and use quotation marks around any exact wording you did borrow. (Here's some advice on the sorts of careful paraphrasing and quoting you need to do to avoid mosaic plagiarism.) You should also be aware that it's possible to be guilty of self-plagiarism, reusing your past work without

acknowledging that fact. This is quite common among students, who may not be aware that they need to get their instructors' permission to adapt work done in one class for another class.

Plagiarism has damaged many careers—sometimes many years after the offending publication. Academics have been reprimanded, even when they footnoted the sources of the wording they used. Students have faced dire consequences just for copying a few sentence fragments. If you feel a little afraid and paranoid after reading this warning, consider it a good thing—and let it motivate you to take extra care to ensure that you are not plagiarizing the work of others.





Many scholars have gotten into trouble for copying phrasings from other people's writing—even when they provided citations. For example, Stephen E. Ambrose, a historian who wrote widely read presidential biographies, was later found to have lifted passages from other writers that he used in several of his books; in his defense, Ambrose said he should have used quotation marks but did footnote the sources—still an academic sin known as *mosaic plagiarism* (Plotz 2002). Pulitzer Prize–winning historian Doris Kearns Goodwin committed a similar kind of plagiarism by swiping numerous phrases and sentences for her book *The Fitzgeralds and the Kennedys* (1991), a scandal that ultimately led to her resigning from the board that awards Pulitzer Prizes and from a gig with *PBS NewsHour* (Anon 2002). *Jim Wallace (Smithsonian Institution), via Flickr / Rhododendrites, via Wikimedia Commons*

Academic disciplines usually have a particular citation style that dictates the format and information included in **in-text citations**—how sources are identified within the text of the paper—as well as in the references listed at the end of the paper. Most social scientific journals, including many sociology journals, use the citations style endorsed by the American Psychological Association (APA style). Books (including academic books) often use Chicago style, which is specified in the *Chicago Manual of Style*, the most popular version of which places citations in footnotes rather than in the main text. APA style is concise, but it leaves out details like authors' first names; Chicago style is elegant and

inconspicuous, but it may annoy academics who don't want to hunt through footnotes to find out which author is being cited.

The style endorsed by the ASA is quite similar to APA style but contains authors' first names, and—in our humble opinion—it just looks better (APA should stand for "ampersands aplenty"). ASA style is used by many of the field's top journals. We'll provide a brief overview, but note that the Purdue University Online Writing Lab also has a helpful guide. For more detailed information, consult the most recent edition of the *ASA Style Guide*.

In ASA style, in-text citations must include the last names of any authors and the year of publication. See **Table A.1** for examples on how to handle most citation situations. One common error that many students make is including a parenthetical in-text citation with an author's last name when that author has already been mentioned in the main text. In that case, you only need to include the year of publication within parentheses right after the author's name (and feel free to use either the full name or just the last name when mentioning cited authors within the main text). Furthermore, you don't have to repeat the same citation multiple times within the same paragraph if it's reasonable to assume, based on context clues, that the information comes from the same source. You can use a single cite when it is first relevant or just at the end of the paragraph—whichever makes more sense.

Another point of style worth flagging is that when a source has more than three authors, ASA style and many other citation styles condense the in-text citation so that it includes only the first listed author's last name and then the Latin term "et al.," which is short for "et alia" (meaning "and others"). You can use "et al." for parenthetical citations, but in the main text, write out all the authors' names, or use the first author's name and "collaborators"—for instance, "Gabriela León-Pérez and her collaborators."

Why does the first author get all the attention? Sometimes—but not always—the first author did more of the work (in terms of data collection, data analysis, or just thinking through the topic) than the other authors. To recognize that larger contribution, the authors might agree that a particular author might be listed as the "first author" even if their last name does not come first in the alphabetical listing. For example, if Gabriela and Victor have coauthored an article, but Gabriela contributed more to the intellectual work, then the in-text citation would be "León-Pérez and Chen." If, however, (1) Gabriela and Victor contributed equally to the article, or (2) Victor contributed more than Gabriela, then the in-text citation would be "Chen and León-Pérez." When the last names are listed alphabetically, as in the latter example, would scenario (1) or scenario (2) apply? The only way of knowing that is if the authors indicated somewhere in their article (such as in the acknowledgments section) how much each scholar contributed to the study.

Table A.1. In-Text Citations

Citation Situation	How to Format It	Example
If you mentioned the author(s) by name in the sentence	Add the year of publication in parentheses after the name(s) of the author(s).	As a consequence, Massey (2002) explains that more unauthorized immigrants have settled rather than returned home after working temporarily in the United States.
If the sentence summarizes or paraphrases the author's ideas, but it doesn't include the author's name	Put the author's last name and year of publication in parentheses at the end of the sentence. Note that in ASA style, there is no comma before the year of publication.	As a consequence, more unauthorized immigrants have settled rather than returned home after working temporarily in the United States (Massey 2002).
If there are two authors	Use the word "and" between the last names.	Mexican immigrant households have lower incomes and higher rates of poverty than the overall immigrant and native-born populations (Zong and Batalova 2018).
If there are three authors	The first time you cite this source, list each authors' last name. In subsequent citations, mention only the first author's last name, followed by "et al." (which means "and others")	Mexican immigrant women's work and family stress are jointly shaped by precarious work arrangements, legal status, and traditional gender ideologies (León-Pérez, Richards, and Non 2021). A prior study found that Mexican immigrant working mothers face additional hardships related to legal status (León-Pérez et al. 2021).
If there are four or more authors	Only mention the first author's last name, followed by "et al."	As above.
If a parenthetical citation includes multiple sources	List the references in alphabetical order and separate them with semicolons. List them in a single logical order throughout the paper—either chronologically or alphabetically (and use the same order consistently throughout the paper.)	*Alphabetically*: (Szeman 2015; Vallas and Prener 2012; Watson 2012) *Chronologically*: (Vallas and Prener 2012; Watson 2012; Szeman 2015).
If you are quoting directly from the text	Add a colon after the year, followed by the page number where the quote can be found (note that there are no spaces between the year, colon, and page number)	As noted by Viruell-Fuentes and her collaborators (2012:2103), "The weight of immigration policies is such that immigrant legal status and citizenship have become central dimensions of stratification in U.S. society."

The end of your paper should include a list of all of the sources that were cited in the text. Typically, you begin this list on a separate page headed with the word "References" or "Works Cited." References should be listed in alphabetical order by the authors' last name. Order multiple works by the same author by their publication year (earliest first). If the author is an organization, use the organization's name in place of a last name. If there is no author, use the first word of the work's title.

The first author's name will always be inverted: list the last name, followed by a comma, followed by the first name (and any middle names or middle initials). In ASA style, the second and subsequent authors are listed with their names not inverted.

How each reference is formatted depends on the type of source. The references for books usually include the following elements:

1. The first and last names of the author(s). Make sure to spell out the first names; don't use initials unless that's the way the name is written in the source.
2. Year of publication
3. Title of the book (in italics)
4. Place of publication (followed by a colon). For cities outside of the United States, provide the country. For U.S. cities, provide the two-letter postal abbreviation for the state (e.g., AK or NY, though you can skip the state for well-known cities such as New York and Chicago).
5. Publisher's name

Here is the format for a book (note the capitalization of titles and names):

> Author 1 Last Name, Author 1 First Name, Author 2 (full first name and full last name, not inverted), and Author 3 (full first name and full last name, not inverted). Year of publication. *Name of Publication*. Publisher's City, State (or City, Country): Publisher's Name.

The references for journal articles should always include the following elements:

1. The first and last names of the author(s)
2. Year of publication
3. Title of the article (in quotation marks)
4. Name of the journal (in italics)
5. Volume and issue numbers. Note that there is no space after the volume number. Put the issue number in parentheses, followed by a colon.
6. Page numbers. Note that there is no space after the colon.

Here is the format for a journal article (note the capitalization of titles and names):

> Author1 Last Name, Author 1 First Name, Author 2 (full first name and full last name, not inverted), and Author 3 (full first name and full last name, not inverted). Year of publication. "Title of Article." *Journal Name* Volume(Issue):Page numbers.

Table A.2 presents examples of how to format common source types in the "References" or "Works Cited" section at the end of the paper.

Table A.2. References in the Works Cited Section

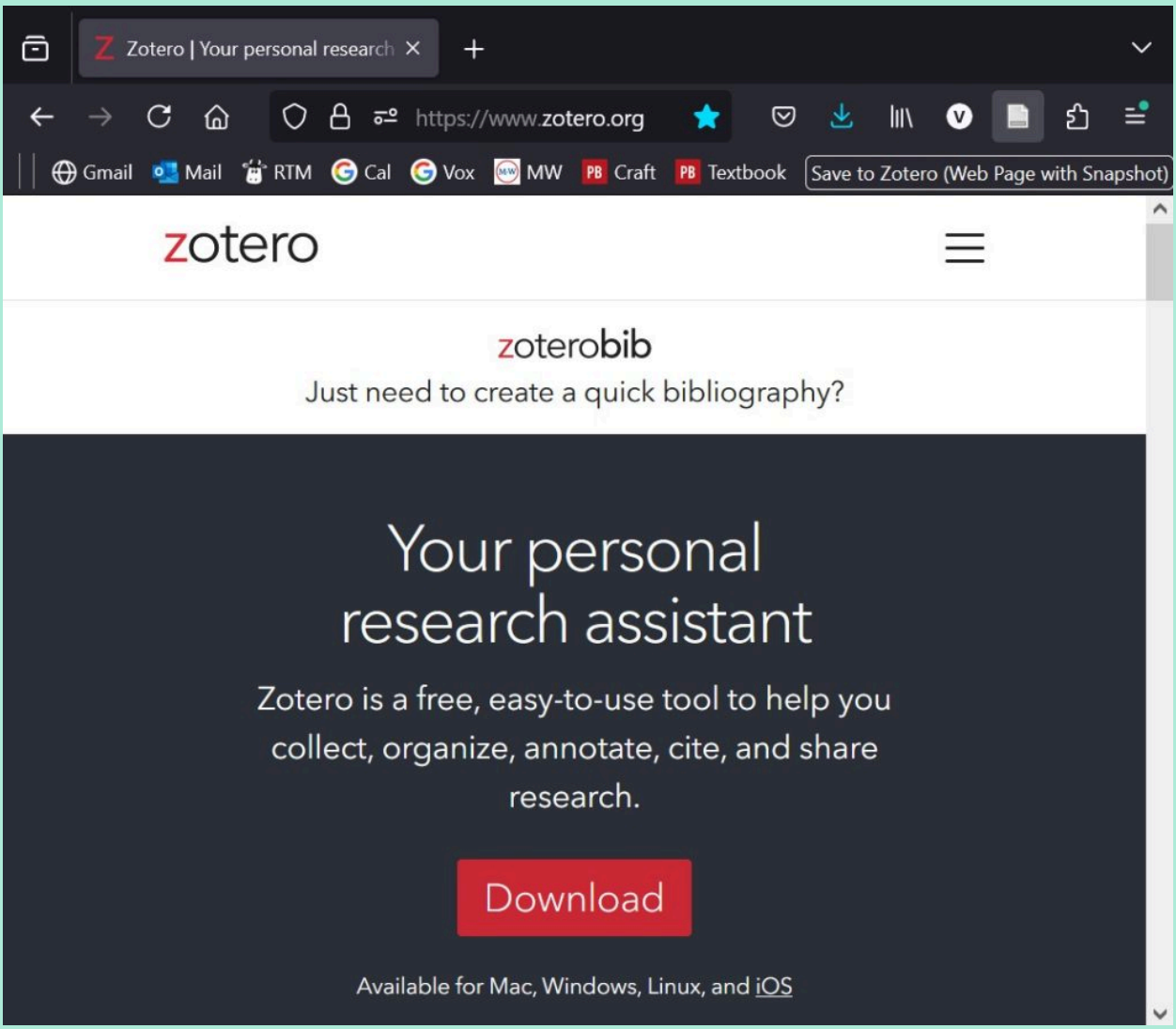
Type of Source	Example
Books, one author	Zuberi, Dan. 2006. *Differences That Matter: Social Policy and the Working Poor in the United States and Canada.* Ithaca, NY: Cornell University Press.
Books, two authors	Newman, Katherine S., and Victor Tan Chen. 2007. *The Missing Class: Portraits of the Near Poor in America.* Boston: Beacon Press.
Books, three or more authors	Massey, Douglas S., Jorge Durand, and Nolan J. Malone. 2002. *Beyond Smoke and Mirrors: Mexican Immigration in an Age of Economic Integration.* New York: Russell Sage Foundation.
Books, edited volumes	Newman, Katherine S., ed. 2010. *Laid Off, Laid Low: Political and Economic Consequences of Employment Insecurity.* New York: Columbia University Press.
Chapter in a book	Jacobs, Elisabeth, and Katherine S. Newman. 2010. "Rising Angst? Change and Stability in Perceptions of Economic Insecurity." Pp. 74–101 in *Laid Off, Laid Low: Political and Economic Consequences of Employment Insecurity*, edited by Katherine S. Newman. New York: Columbia University Press.
Journal article, one author	Hondagneu-Sotelo, Pierrette. 1992. "Overcoming Patriarchal Constraints: The Reconstruction of Gender Relations among Mexican Migrant Women and Men." *Gender and Society* 6(3):393–415.
Journal article, two authors	Portes, Alejandro, and Min Zhou. 1993. "The New Second Generation: Segmented Assimilation and Its Variants." *The Annals of the American Academy of Political and Social Science* 530:74–96.
Journal article, three or more authors	Doody, Sean, Victor Tan Chen, and Jesse Goldstein. 2016. "Varieties of Entrepreneurial Capitalism: The Culture of Entrepreneurship and Structural Inequalities of Work and Business Creation." *Sociology Compass* 10(10):858–76.
Articles from magazines and newspapers	Reeves, Richard V., and Isabel V. Sawhill. 2015. "Men's Lib!" *New York Times*, November 14, SR1. Retrieved August 29, 2018 (https://www.nytimes.com/2015/11/15/opinion/sunday/mens-lib.html).
Authored by organizations	U.S. Census Bureau. 2018. *POV-13. Related Children by Number of Working Family Members and Family Structure.* Washington, DC: U.S. Census Bureau. If your source is a report, then the title of the report should be put in quotation marks rather than italicized: U.S. Department of Health and Human Services. 2012. "2012 Annual Report of the U.S. Organ Procurement and Transplantation Network and the Scientific Registry of Transplant Recipients." Washington, DC.
Information posted on web pages	"Department of Sociology." 2024. Virginia Commonwealth University Department of Sociology. Retrieved August 18, 2024 (https://sociology.vcu.edu).

Once you start collecting a substantial number of references for your manuscript, you will quickly realize how tedious it is to look up the authors, titles, publication years, and other bibliographic information for every article and book you are citing. Things will get even worse when you organize your reference list at the end. By then, your head will probably be throbbing in pain as you frantically try to remember whether ASA style has a space after the colon before a journal article's page numbers (it doesn't) or whether the volume number goes before the issue number (it does).

Fortunately, no-cost automated technologies exist to help you. For example, Zotero is free and open-source software

that will collect bibliographic information and format it properly for you. (Other applications like EndNote, Mendeley, and RefWorks are similar, but unlike Zotero, they are not open-source and charge money to use all or some features.) Zotero can be downloaded as an app for desktop computers, tablets, and smartphones at zotero.org. Once you have installed Zotero's browser extension or mobile app, you can usually obtain the necessary bibliographic information for a particular source by going to an associated webpage—a journal article page or a library catalog entry, for instance. With the click of a button, Zotero will pull details like the article title, journal name, and publication date, which the software scrapes from the webpage's metadata (the sidebar **Using Zotero to Manage Your References** walks you through this process). What's more, Zotero automatically generates in-text citations and reference lists in word processors like Google Docs for the sources you have collected. Finally, you can store, read, and annotate the PDF files for your sources right in the Zotero app, which offers a free basic plan that lets you keep these files in the cloud so that you (and, optionally, colleagues you are working with) can access them from any device.

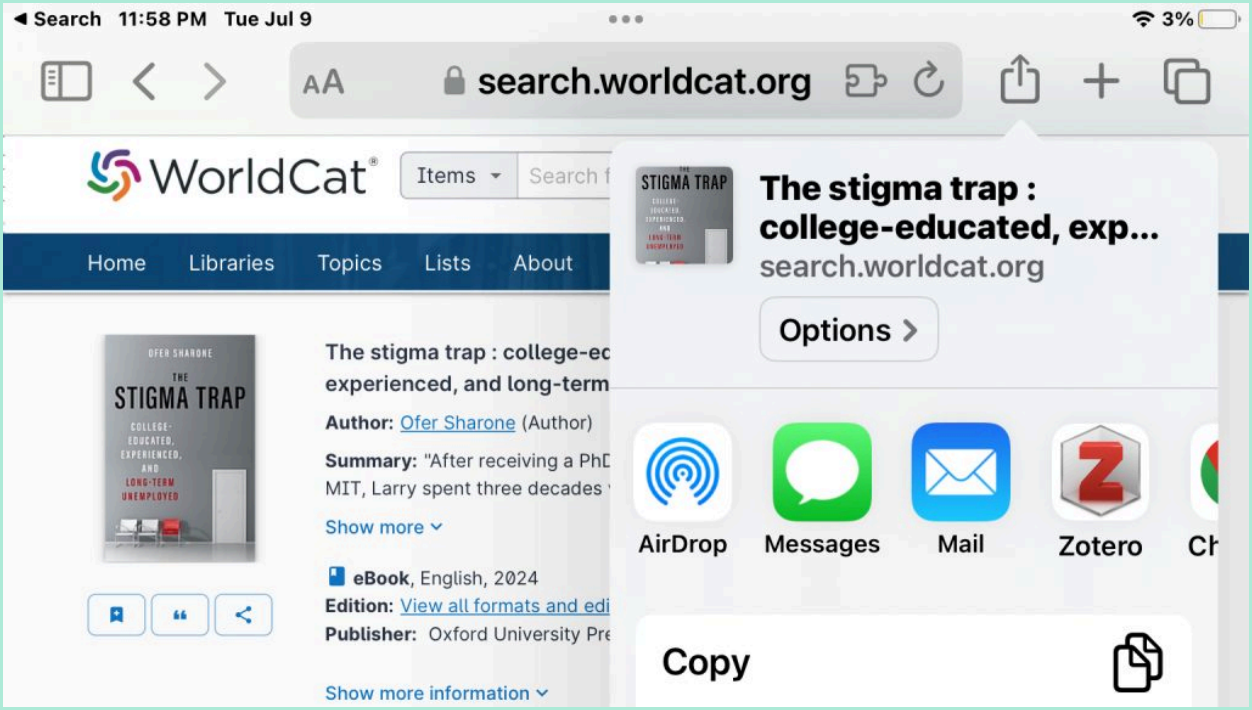
Deeper Dive: Using Zotero to Manage Your References



Zotero is a free and open-source application that helps organize your references and generates properly formatted citations and reference lists for you. It can be downloaded from the website shown here, zotero.org. Note the highlighted Zotero button to the right of the address bar, which appears after you install the Zotero browser extension.

The following instructions cover the basics of Zotero, free and open-source software widely used by academics to manage the citations and reference lists in their manuscripts. For more detailed instructions, visit Zotero's Documentation page and its Quick Start Guide.

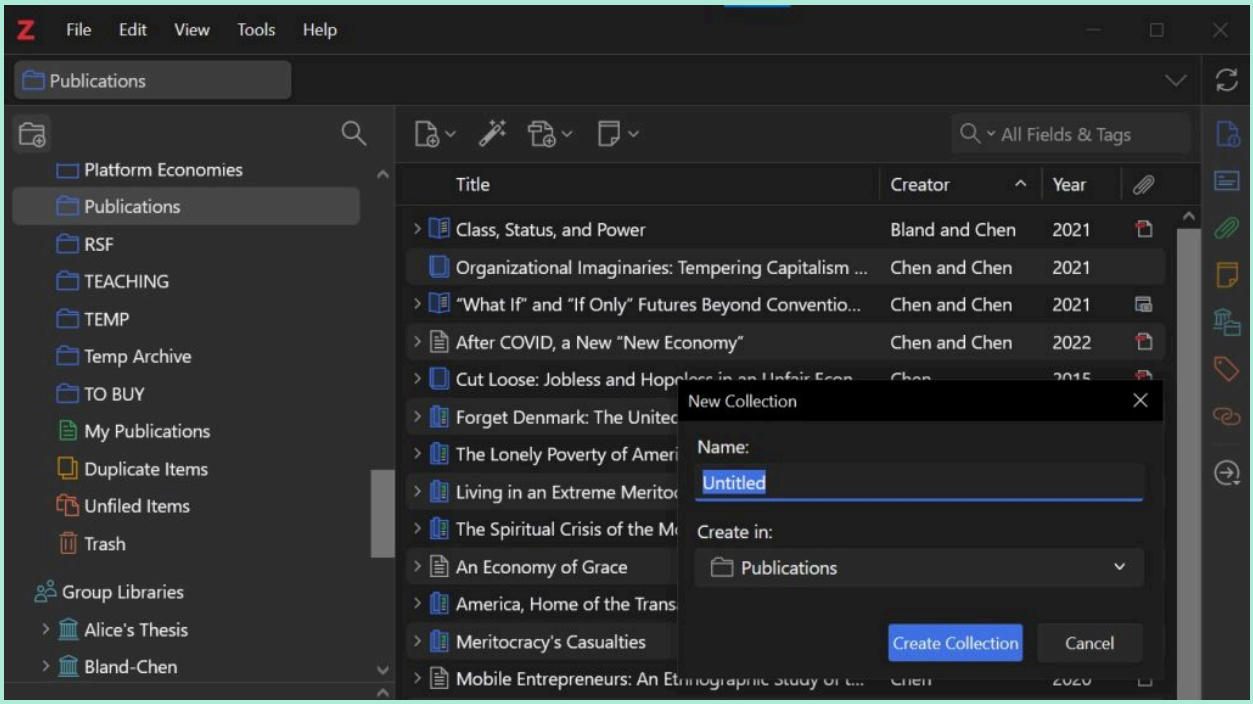
1. To set up Zotero on your computer, go to zotero.org and download both the Zotero desktop application and the Zotero Connector extension for your particular browser (Chrome and Firefox are supported). If you want, you can download and install the Zotero app on your Android or iOS mobile device as well.
2. Install both the desktop application and the browser extension. After you install the extension, a Zotero button will appear to the right of your browser's address bar. (If you have multiple extensions installed, you may have to click on the extensions link to the right of the browser's address bar to see this additional option.) Run the desktop application and go to **Edit > Settings**. Click on **Sync** on the left. Log into your Zotero account; if you haven't already set up a Zotero account, click on **Create Account** or go to zotero.org to register. Keep the desktop application running in the background and then open your browser. (You can do some of the following procedures without the desktop application running, but it is not recommended.) If you are using Zotero on a phone or tablet, go to the app's settings (the gear icon within the app) and log into your Zotero account there. Then open whatever browser you wish on your mobile device.
3. To add a source to your Zotero library, visit any webpage *relevant to that source* in your browser. For example, to add a book to your library, you can visit its page on Amazon, Bookshop.org, Google Books, any library catalog (the combined catalog WorldCat is a good place to start), or the publisher's website. (Note that academic publishers often fail to supply sufficient metadata for their books, so the other options listed may be a better bet.) For a journal article, visit that article's page on the journal's website; you can just do a web search using the article's title and its author's name, or if you know the article's **digital object identifier** (DOI), you can find it easily by going to the following URL in your browser: `https://doi.org/[DOI]` (with the article's DOI replacing `[DOI]` in the address bar).
4. If you are using Zotero on your computer, click on the Zotero button to the right of your browser's address bar. This button should have changed to an icon representing the type of content associated with the webpage you are visiting at the moment in your browser (a book, journal article, etc.). Clicking the button will automatically save the source's bibliographic information to whatever Zotero library collection or subcollection (i.e., folder or subfolder) is currently open in your Zotero desktop application. (Remember to have the desktop application running in the background whenever you click on the Zotero button in your browser.) If you are using Zotero on a phone or tablet, choose the **Share** option in the browser menu and select the Zotero app. Tap on **Save to Zotero** (make sure that you have logged into your Zotero account on the app before doing this).



On a mobile device, you can share any webpage you are browsing with the Zotero app. For instance, if you look up a book in a library catalog like WorldCat and then share the link with the Zotero app, it will automatically store the bibliographic information for that book in your Zotero library.

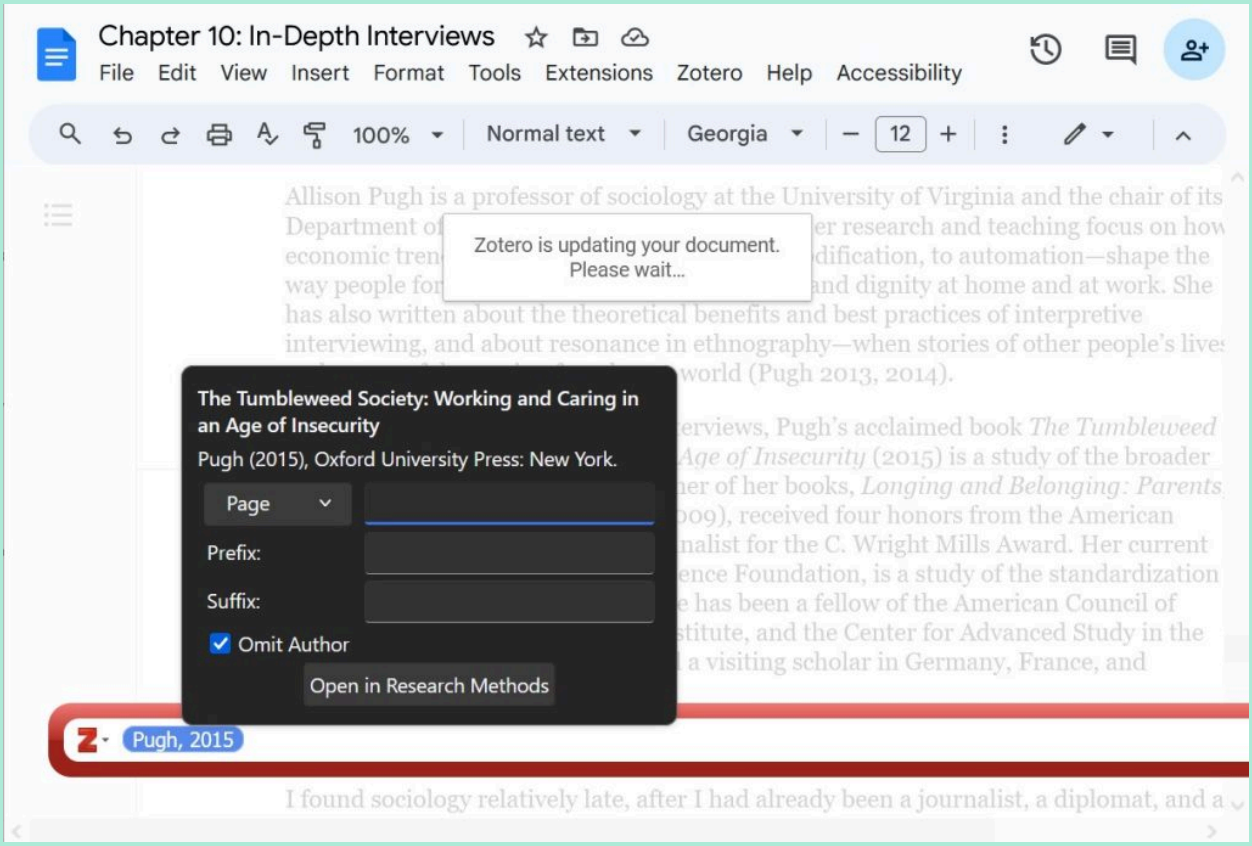
The following are some additional tips for using Zotero.

Collections and subcollections. You can create your own collections and subcollections in your Zotero library. In the desktop application, click on the folder icon in the top-left corner, or right-click on an existing folder and select **New Collection** or **New Subcollection**; in the mobile app, click on the plus button. When saving a source's bibliographic information to your library, you can change the selected collection or subcollection in the menu that appears. To view all your library's folders, choose the **More** option at the bottom. To reorganize these folders, click and drag a collection into another collection (in the desktop application) or tap on a collection's name and then tap **Edit** to change its parent folder (in the mobile app).



You can create collections in your Zotero library to organize your sources. In the desktop application, click on the folder icon in the upper-left corner and then click and drag sources into the newly created folder. If you wish, you can also create subcollections nested within parent collections. Note that any Zotero Group Libraries that have been shared with you (as described at the end of this section) are listed in the left pane below the personal library collections.

Editing and manually inputting references. Sometimes, the reference information you find on webpages for a particular source will be incorrect or incomplete. You can edit whatever information is associated with that source within the desktop application or mobile app. You can also input a source manually. In the desktop application, click on the document icon with a plus sign at the top of the middle pane; in the mobile app, make sure you are in the correct collection or subcollection and then tap the plus button. Choose the appropriate source type (**Book**, **Journal Article**, etc.) and then enter the reference information.



In a word processor like Google Docs, you can insert an in-text citation for any source in your Zotero library and also specify whether to include or omit the name of the author (useful if you already mentioned their name earlier in the same sentence, as shown here). Note that the Zotero menu appears in the Google Docs menu bar at top after you install the Zotero browser extension.

Zotero in your word processor. Once you have added sources to your Zotero library, you can easily create citations and reference lists in your word processor. More detailed instructions for integrating Zotero into various word processors (including Microsoft Word and the open-source LibreOffice Writer) are available on Zotero's website, but here are basic instructions for Google Docs:

1. Keep the Zotero desktop application running in the background and open the same browser you were using that has the Zotero extension installed. (You can do some of the following procedures without the desktop application running, but it is not recommended.) When you open Google Docs, a **Zotero** menu should now appear in the top menu.
2. Position your cursor wherever in the document you wish to insert an in-text citation. Then select **Zotero > Add/edit citation**. You will need to log into your Zotero account and then choose the citation style you wish to use (for instance, **American Sociological Association (ASA) 6th/7th edition**).
3. A Zotero search bar will appear. Search for the source you just added to your Zotero library using whatever title or author keywords you wish. Select the correct source from the list, which will automatically populate as you type. You can keep adding sources if you want to cite multiple sources

within a single parenthetical citation. If the author's name has already been mentioned in the text and you just want to include the publication year in your citation, you should click on that source in the search bar and then check the box **Omit Author**. When you are done adding sources, press the enter key from anywhere within the Zotero search bar. The citation will be inserted into the text of your Google Doc.

4. To add an automatically generated and updated reference list to the end of your document, first position your cursor wherever in the document you wish to place the reference list. Then select **Zotero > Add/edit bibliography**. The full bibliographic information for any in-text citations within that Google Doc will be pasted into that location in the document. You can revise the bibliographic information for any of these sources at any time in the Zotero desktop application or mobile app, but for those revisions to appear in your reference list, you may need to select **Zotero > Refresh** from within your Google Doc. If you prefer to create the reference list manually, you can obtain the properly formatted bibliographic information for any source by right-clicking on the source in the Zotero desktop application and choosing **Create Bibliography from Item**. Choose the option **Copy to Clipboard**, press **OK**, and then paste (CTRL-V on a PC) the text anywhere you wish.

Zotero Groups. If you are working with other researchers, you can create a Zotero Group to share your references with your colleagues and maintain a single up-to-date library of sources. Log into your account at zotero.org, click on your profile, and click on **Groups** in the top menu. Then click on **Create a New Group** at top. After you set up the group and invite other people to join it, open the Zotero desktop application or mobile app. Make sure you are logged into whatever Zotero account (under **Edit > Settings**) is associated with the Zotero Group you created. The new group should appear as a separate library below your personal library (**My Library**); if you don't see it, restart the application or check your app's sync settings. Once the Zotero Group is set up, you and your colleagues can add references to it by following the earlier instructions. Note that any changes you make to sources stored in your personal library will not be made to the same sources stored in a Zotero Group, and vice versa.

Key Takeaways

1. Reports for public consumption usually contain fewer details than reports for scholarly consumption.
2. Keep your role and obligations as a social scientist in mind as you write up research reports.
3. Plagiarism is among the most egregious transgressions a scholar can commit, and it includes situations in which you do cite the sources you are borrowing from but follow their original wording too closely (always use quotation marks or paraphrase properly!).

Exercises

1. Imagine that you've been tasked with sharing the results of some of your research with your parents. What details would you be sure to include? What details might you leave out, and why?
2. Have a discussion with a few of your peers about plagiarism. How do you each define the term? What strategies do you employ to ensure that you avoid committing plagiarism?

A.4. Reaching Your Intended Audience



Sadly, researchers aren't rock stars. If you want a particular audience to learn about your work, you often need to spread the word yourself—by posting to social media and networking with other scholars through professional associations, among other things. To reach policymakers or mainstream audiences, you will need to go further: writing op-eds, landing media interviews, and giving talks for think tanks and other influential groups. *Vishnu R Nair, via Pexels*

Learning Objectives

Identify the ways that you can disseminate your research to both scholarly and non-scholarly audiences.

Writing up results from your research does not necessarily mean others will take notice of it. For academics, the general rule of thumb is that people will *not* take notice unless you help and encourage them to do so. Once your work is out there, you need to make sure it reaches its intended audience.

Where your audience is should be fairly obvious once you've determined who you'd like your audience to be. You

know where your research participants are because you've studied them. We've already talked about reaching scholarly audiences by presenting at professional conferences or publishing in academic journals. You can also look to your own campus: universities and colleges often run their own research **symposiums** (a panel or series of panels devoted to a particular issue) or conferences that feature student research talks, and any campus event could be a possible occasion for sharing relevant study findings. An easy way to reach other scholars who study your topic is to write something for the newsletter or blog published by a professional organization (or the relevant section or network of that larger organization). This is an often-overlooked source for sharing research findings in brief form with potentially interested scholars.

What about if you think your work has broader appeal and you want to reach nonacademics? If that's the case, you need to rethink the way you write up your results. Think about what your intended audience is likely to want to hear about. What portions of the research do you feel are crucial to share, regardless of the audience? And what portions do you think might bore nonacademics? These questions should guide you in writing up your report. In fact, some outlets answer these questions for you—a particular news publication's rules of style, presentation, and length will dictate the shape of your written report, for example (see the sidebar **Writing Op-eds**).

Whoever your audience is, don't forget what it is that you are reporting: social scientific evidence. Take seriously your role as a sociologist and your obligation to advance scientific knowledge. Present your findings as clearly and as honestly as you possibly can. Give appropriate credit to the scholars who have come before you, even while you raise thoughtful questions about their work. And aim to engage your readers in a discussion about your work and about avenues for further inquiry. Even if you won't ever meet your readers face-to-face, imagine what they might ask you upon reading your report, and imagine your response. You'd be wise to provide some of those details in your written report to anticipate any concerns or questions.

In terms of reaching policymakers, your state and federal representatives in theory should be available to hear a constituent speak on matters of policy interest. That said, it's usually better to reach them through intermediaries—either media outlets or organizations that work with policymakers. For example, the Washington Center for Equitable Growth funds researchers who study economic issues, and the center works with the recipients of its grants to produce working papers and policy briefs related to their research, sharing them with the policymakers who are part of the center's network. Perhaps you're already aware of organizations that do work in an area related to your research topic, but if not, a simple web search should help you identify possible organizational audiences for your work.



Many sociologists dream that their work will someday shape high-level policy decisions. Until the White House establishes a Council of Sociological Advisers to balance out its Council of Economic Advisers, that sort of influence is unlikely. Yet there are still things we can do as sociologists to disseminate our knowledge widely and possibly reach policymakers—among other things, by connecting with think tanks that regularly advise elected officials and other key decision-makers. *Aaron Kittredge, via Pexels*.

Disseminating your findings to the public more generally could take any number of forms: anything from posts on social media to an article in a news outlet. These days, academics are highly active on X (formerly Twitter), Facebook, and LinkedIn, where they spread news about their research or comment on other studies (thereby drawing attention to their own work). The sidebar **Writing Op-eds** describes how to pitch commentary pieces to news outlets—which will publicize pieces you write for them on social media, too. Some organizations also work with academics to write up their research in a form accessible to mainstream readers or policymakers. For example, the Scholars Strategy Network, a politically progressive organization with chapters across the country, will help pitch academics' commentary pieces to mainstream publications. *The Conversation*, a nonpartisan news outlet, specializes in helping academics translate their work into the form of articles published on *The Conversation*'s website. (Like this textbook, *The Conversation* publishes its content under a Creative Commons license, meaning that other publications can republish it with some restrictions.)

Writing Op-eds, by Victor Tan Chen



Victor has written op-eds about his research for the mainstream news site *The Atlantic*, such as this featured article about unemployed working-class white workers (Chen 2016). *Atlantic Media*

Many sociologists think our work would be helpful for policymakers and the broader public to understand, but we struggle to get our views heard outside a small circle of fellow researchers. While we're waiting for a CNN or *New York Times* reporter to stumble across our research and call us, we can also consider writing nonacademic pieces that make the findings of our research accessible to ordinary readers. *Atlantic*

It's important to emphasize that writing for public consumption is very different from the writing you do for an academic paper. You should generally think of them as separate projects, because the language you use in one will be very different from what you use in another. In fact, even if you haven't done empirical research on a particular topic, you might find news outlets willing to publish your writing so long as you have an interesting take on the research that other people have done. As we described in **Chapter 2: Using Sociology in Everyday Life**, students who are seeking careers outside academia can find success as journalists or public relations staff who translate research for nonacademic audiences in this way.

The sorts of essays that academics typically write for news publications are opinion or commentary pieces called **op-eds** (which stands for "opposite the editorial page," referring to where these guest essays were traditionally placed in a newspaper). Generally, op-eds have a clear argument that can be summarized in a thesis statement of one or a couple sentences—which means that the version of a particularly complex piece of research for public consumption will often need to be broken up into multiple op-eds. You should be aware that it is okay to include your personal perspective in an op-ed. Most publications prefer that you use first-person voice, for example, and connecting your research to your own life can be engaging for readers. In fact,

in addition to traditional op-eds, publications often accept articles that are entirely personal essays but connected to a news topic.

For better or worse, many prominent publications want a strong opinion expressed in an op-ed. In other words, they want you to have a forceful, provocative argument—one that doesn't dance around with all the nuances and contradictions of your findings, but clearly states the driving point and its broader implications. (As an editor once told me, an op-ed is a "made-up mind.") It is a judgment call you will need to make, whether you are comfortable with presenting your research in this way, and we would strongly advise against twisting your research findings to fit a particularly compelling narrative. That said, a certain degree of simplification always has to go into presenting your work to a nonspecialist audience, and it is a helpful exercise to learn how to boil down your findings into a concise, powerful argument.

Identifying publications to query. Which publication should you try to write for? You can pitch almost any publication, but you will want to consider its readership and whether that's the sort of audience you want to reach. Unless you're an astounding writer or have a research finding that blows people away, you probably don't want to bother with the largest and most prestigious publications; they tend to look dimly upon newbie researchers without an established track record. You also should note that publishing with highly political outlets may (rightly or wrongly) associate you with their ideologies and partisan leanings, meaning other scholars may look at you and your work differently. In general, it's a good idea to query publications that you personally enjoy reading, which means your voice or interests as a writer will probably match their own. For folks new to op-ed writing, smaller publications may be a good place to get constructive (or any) feedback on your writing style, since their editors are not as deluged with pitches; for instance, since 2001, I've worked on *In The Fray*, an online magazine devoted to personal stories on global issues, where I've mentored young writers who are just learning the craft.

Querying publications. How should you go about reaching out to ("querying" or "pitching") publications? First, you should get to know the publication you wish to target by reading through past articles they've written on your topic. On the one hand, if the subject of your piece is too far astray from the sorts of articles that the publication typically publishes—in terms of either the substance or the style of your writing—the editors may reject your submission right off the bat. On the other hand, if your op-ed is too similar to what they've recently published, you might be rejected for that reason, too. If you think the publication is a good fit, then read their submission guidelines carefully. Every publication has such guidelines for freelancers (writers who aren't staff), though you may have to do some digging online to find them.

The submission guidelines will tell you whom to contact about your article, what to say in your submission, and when (if ever) to expect a reply. Sometimes, editors will only accept a full draft of an op-ed in your pitch. Newspaper opinion pages like those for the *New York Times* and *Washington Post* often do this. Many other publications, however, prefer that you send them just a **pitch**—a brief description of what you propose to write about—rather than a completed draft. As noted in the *Guardian*'s submission guidelines, editors at some publications don't have time to read full drafts and would rather work with you from the get-go to ensure that your piece meets their expectations.

Writing a pitch. The pitch should be no more than a paragraph long—several sentences that concisely summarize what you want to argue, perhaps with one or two examples or pieces of evidence to support your claims (with links to back you up). Providing a news "peg" or news "hook"—a description of how your article is related to hot topics currently in the news—can be helpful, too, but as *Slate* notes in their submission guidelines: "Avoid vague, one- or two-sentence pitches on a topic, as in, 'I would like to write about X because of Y news peg.' A short paragraph that captures your argument will be more effective." Tailor the focus of your

piece to the publication's audience, if possible; if it is a national publication, you don't want to focus too much on local issues unless you can make a strong case they have national significance.

Regardless of whether you are including a full draft in your submission, it's important to have a provocative and tightly written pitch paragraph in the email you send to editors. Otherwise, they may not bother to read your full draft. Likewise, when you send your query (which is typically done by email), the subject line should not be a generic subject (e.g., "Op-ed Submission"), but rather it should describe your thesis succinctly (like a headline would) or at least mention the specific topic. Along with your pitch paragraph, include a line or two about who you are and your qualifications for writing this piece; you might want to include links to past pieces you've written if they're particularly notable and they're also similar to what the publication publishes. If you're including a full draft of your op-ed, you can usually just paste it into the text of the email rather than including an attachment; check the publication's guidelines.

Don't query more than one publication at a time. As with academic journals, wait until the publication rejects your piece before moving on to another publication.

Writing the op-ed. The most important rule in writing a nonacademic essay is to *avoid jargon*. Obviously, this includes social-scientific terminology, but also consider that for more mainstream publications, you will want to avoid any language or phrasings that are used largely in the academy and not widely adopted elsewhere. For publications geared to broad audiences, think about what a parent, grandparent, aunt/uncle, or sibling, etc., would be able to relate to. (My pet peeve is the constant use of partially parenthetical terms by academics—"on (hu)man bondage" and its ilk—which no one outside the academy seems to find as witty.) I would also recommend taking it easy on the use of acronyms, which social scientists tend to adore. Overuse of them can turn the text into an alphabet soup and make your readers' eyes glaze over. On second reference, you might choose to refer to something by part of its name or by another descriptor rather than using an acronym. For example, refer to "the university" rather than "VCU," and refer to "this theory" rather than "RIT" (for "relational inequality theory") or "WST" (for "world-systems theory") or whatever arbitrary agglomeration of letters you use for your pet theory.

Especially for an op-ed, the first paragraph (the "lede" or "lead") is typically a brief statement of your thesis. If it is not, then a paragraph near the top of the piece should briefly summarize your argument and describe why a reader should care about it (what's sometimes called a "nutgraf," since it says why your piece matters in a nutshell). Generally, op-ed pieces should be between 500 and 1,000 words (check the publication's website, though). This means you need to keep your take on the research short and sweet. You can include some empirical data as vignettes or statistics, but don't overdo it. For statistics, you will just want to flag the most pertinent numbers, and then more generally describe any relevant trends or patterns in the data. For qualitative data, you typically want to provide quotes from a single respondent or observations from a single event to illustrate a single point, with the expectation that additional data exists to back you up. For the most part, use shorter quotes from interviews, and include a brief description of the speaker (typically, noting their age, gender, race, and where they live). Publications have their own rules about whether pseudonyms should be used (or if unnamed respondents are even allowed); check with an editor.

Unlike for academic papers, the rules for sourcing your information in an op-ed are looser. Don't include academic-style citations (e.g., last names and page numbers), but do reference any person you quote, or the source of important information, in your article's prose ("According to so-and-so"). Mentioning lots of academics or a lot of studies will make your piece sound too academic, so if you're writing an online piece, you might want to instead link to that research rather than mentioning it in the text.

Note that you do not control the headline of your piece, which is often written by editors after you sign off on your draft. Furthermore, you normally will need to defer to the editor about the final version of the piece,

including changes to specific phrasings within it—what are known as *line edits*. (Nonacademic editors typically exert much more control over your writing than academic editors do.) That said, don't be afraid to stand up for yourself on important points. At the end of the day, you want to publish a piece that accurately represents your research.

Following up. After you send your query, you will need to wait patiently for a reply. Publications often list the expected wait time in their submissions guidelines; typically, it is one or two weeks. If you don't hear back after a week (or whatever time they specify in their guidelines), you can send one follow-up email, but don't pursue it beyond that, and don't call them. Once the expected wait time passes, you can move on to pitch your piece to another publication.

Larger publications often pay their writers, though the amounts are usually quite small for op-eds. Importantly, op-eds are written "on spec," meaning that they can be rejected at any time, and there is no guarantee of payment until the piece is published. You will have to decide whether the lack of pay justifies all the work you put into crafting an op-ed. That said, there is a great deal of satisfaction in bringing sociological ideas to the broader public. At the very least, you will finally be able to explain to your friends and family what it is you do exactly.

Key Takeaway

There are many ways that you as an individual scholar can reach scholarly and non-scholarly audiences—from contributing to a professional organization's newsletters, to posting on social media, to writing op-eds for news publications.

Exercises

1. What additional potential audiences for your research, aside from those already mentioned, can you identify? How might you reach those audiences?
2. Do you agree or disagree with the assertion that researchers who conduct high-quality research have a duty to share their findings with others? Explain.

A.5. References

Becker, Howard S. 2020. *Writing for Social Scientists: How to Start and Finish Your Thesis, Book, or Article*. 3rd ed. Chicago: University of Chicago Press.

Chen, Victor Tan. 2016. "All Hollowed Out." *The Atlantic*, January 16.

Davis, Daniel, and Amy Binder. 2019. "Industry, Firm, Job Title: The Layered Nature of Early-Career Advantage for Graduates of Elite Private Universities." *Socius* 5:2378023119859711. doi: 10.1177/2378023119859711.

"Doris Kearns Goodwin Leaves Pulitzer Prize Board." *Wall Street Journal*, May 31.

Goodwin, Doris Kearns. 1991. *The Fitzgeralds and the Kennedys: An American Saga*. New York: St. Martin's.

Johnson, William A., Jr., Richard P. Rettig, Greg M. Scott, and Stephen M. Garrison. 2010. *The Sociology Student Writer's Manual*. 6th ed. Upper Saddle River, NJ: Prentice Hall.

Moskovitz, Cary, and Lynn Smith-Lovin. 2017. *Writing in Sociology: A Brief Guide*.

Plotz, David. 2002. "The Plagiarist." *Slate*, January 11.

Population Reference Bureau. 2003. *Connecting People to Useful Information: Guidelines for Effective Data Presentations*. Washington, DC: Population Reference Bureau.

Reczek, Corinne, and Debra Umberson. 2012. "Gender, Health Behavior, and Intimate Relationships: Lesbian, Gay, and Straight Contexts." *Social Science and Medicine* 74(11):1783–90. doi: 10.1016/j.socscimed.2011.11.011.

Schneider, Daniel, Orestes P. Hastings, and Joe LaBriola. 2018. "Income Inequality and Class Divides in Parental Investments." *American Sociological Review* 83(3):475–507. doi: 10.1177/0003122418772034.

South, Scott J., and Lei Lei. 2021. "Why Are Fewer Young Adults Having Casual Sex?" *Socius* 7:2378023121996854. doi: 10.1177/2378023121996854.

Umberson, Debra. 2015. "One Benefit to Same-Sex Marriage That Nobody Is Talking About." *HuffPost*, May 7.

White, Lynn. 2005. "Writes of Passage: Writing an Empirical Journal Article." *Journal of Marriage and Family* 67(4):791–98. doi: 10.1111/j.1741-3737.2005.00175.x.

Glossary

abductive approach

An approach to empirical investigation in which researchers apply a particular theory to the social context they are examining and then look for deviations from that theory. (Also referred to with the terms *abduction* or *abductive analysis*.)

academic journals

Periodical publications that contain scholarly papers focused on the topic or topics that the journal covers.

academic tenure

A form of job security granted to senior professors who are distinguished as experts in their field. Tenure makes it difficult to fire a professor and is meant to protect their ability to study and write about controversial topics.

acquiescence bias

Bias that occurs when people tend to say "yes" to whatever the researcher asks, even when doing so contradicts previous answers. (Also known as *yea-saying*.)

adjunct faculty

Part-time instructors or lecturers at a college or university who do not have long-term contracts with the university, but who are typically paid for each class they teach. (Colloquially known as *adjuncts*.)

advocacy groups

Organizations that engage in public communications and/or lobbying to influence the views and decisions of important audiences, including policymakers and the general public. (Also known as *interest groups* or *pressure groups*.)

analytic field notes

Field notes that attempt to explain or comment on observations in the field, rather than just describing them.

anecdotal evidence

Data based on people's personal experiences that has not been collected systematically.

anonymity

When no identifying information can be linked to the specific individuals participating in a study, even by the researchers themselves.

antipositivism

A variety of theoretical approaches that critique positivism and its belief in the possibility of arriving at the truth of reality.

applied research

A type of research that applies scientific knowledge to a practical problem.

archival searching

Expanding the coverage of a literature review by identifying any relevant work that (1) is cited in the sources already gathered (as indicated in their reference lists), or (2) cites those sources (as indicated in scholarly databases).

attribute coding

A specialized type of descriptive coding in which the researcher flags relevant characteristics of the case that was observed, such as demographic information for specific interviewees. Attribute coding helps the researcher describe the study sample and allows them to see potential patterns—for instance, how people with certain characteristics responded to interview questions.

attributes

Characteristics of a variable representing its different possible values or categories. For instance, "42" can be an attribute of the variable *age*, and Islam can be an attribute of the variable *religious affiliation*.

attrition bias

Bias that occurs when the participants who leave a study tend to come from particular subgroups, thereby undermining the representativeness of the study's sample. (Also called *mortality bias*, given how test subjects in a classic scientific lab experiment would die out or otherwise leave the study.)

audit

A method used to evaluate the rigor of qualitative research by tracing each point made in the published study back to the original data that was its basis.

authenticity

A criterion for assessing the rigor of qualitative research that focuses on the extent to which researchers capture the multiple perspectives of their research participants. *Fairness* is a dimension of authenticity.

authority

A socially defined source of knowledge that might shape our beliefs about what is true and what is not true.

axial coding

A type of explanatory coding that involves arranging codes to describe causal relationships or hypotheses.

basic research

Research conducted for its own sake. (Also known as *pure science*.)

bias

A systematic error that may make research findings inaccurate in some way. Note that the term "bias" in this context does not just refer to the researcher's personal biases, but to anything that causes a study's results to fail to truthfully represent reality.

bidirectional

When the causality in a relationship runs in both directions—that is, when changes in the first variable cause changes in the second variable, *and* changes in the second cause changes in the first.

big data

Large datasets that typically have been created through automated processes (such as data generated by online platforms) and require sophisticated data-processing techniques to analyze.

blockquote

An extended quotation that is indented (typically by a half-inch) to highlight it in the text.

bystander observation

A type of ethnographic observation in which researchers choose not to get involved in the activities or organizations they are studying, typically with the goal of being more impartial in their assessments of what they observe. (Also known as *direct observation*.)

cases

Members of the sample that the researcher has gathered data on, such as the individual interviewees or organizations being studied.

categorical variables

Variables measured at the two lower levels of measurement, *nominal* and *ordinal*.

causal inference

The process of testing, based on empirical evidence, whether changes in one variable truly cause changes in another variable.

causal mechanism

The specific process or pathway by which one concept affects another. (Also referred to as a *mediating concept*, *mediating variable*, or *linking concept*.)

causal stories

causal story

A theory of how exactly changes in one concept lead to changes in another concept. (Also referred as an *explanatory story* or just a "story.")

certificates of confidentiality

In the United States, a document issued by the Department of Health and Human Services to protect the privacy of a research study's participants. Typically, researchers can be compelled by court orders to release or discuss the data they collect, but certificates of confidentiality (which researchers must apply for) prevent them from having to disclose certain information to anyone not connected to the research.

classical experimental design

An experimental design with a number of features designed to ensure the experiment's internal validity. (Also called a *true experimental design*.) Study participants are randomly assigned to an experimental group or a control group, and pretests and posttests (measurements of the dependent variable) are conducted before and after the introduction of the stimulus or treatment (for the experimental group) or any placebo (for the control group).

clinical sociology

The use of sociological knowledge to improve people's quality of life in a healthcare setting.

closed-ended questions

Interview questions that provide respondents with a limited set of options for their responses. (Also referred to as *close-ended questions* or *closed questions*.) In in-depth interviews, closed-ended questions are often yes/no questions or can be answered with a word or two. In surveys, closed-ended questions require respondents to choose from two or more response categories.

cluster sampling

A sampling technique in which a researcher begins by sampling groups (or *clusters*) of population elements and then selects elements from within those groups.

clusters

In cluster sampling, the groups within the sample that the researcher identifies to systematically draw elements from those groups rather than from the sample as a whole.

code

In qualitative research, a shorthand representation of some more complex set of issues or ideas. It typically takes the form of a label (a word or phrase) that tells us the theme or argument evident within a portion of text. In survey research, codes refer to the numerical values assigned to each survey question's response options to facilitate data analysis.

code-switch

To move between different linguistic or cultural styles.

codebook

For quantitative research, a document that describes the content and structure of a dataset. It typically contains information about the variable names for each survey question and the numerical values assigned to each response option. For qualitative research, codebooks describe the criteria that the researchers used when deciding how to categorize and label their data.

coding

In qualitative analysis, the process of identifying relevant themes through reading and review of the relevant data. In survey research, coding refers more narrowly to the process of assigning numerical values to each survey question's response options so that the respondents' responses can be recorded in a quantitative dataset.

coding memos

Memos that researchers write during the coding process that present their preliminary ideas and hunches about the data being reviewed, often with a focus on a particular set of related codes. (Also called *analytic memos*.)

coding unit

The specific portions of the text that are being labeled with a code in an analysis of interview transcripts or field notes. A coding unit can range from a single phrase to multiple paragraphs. Note that coding units may also take the form of images and other nontextual materials that are being coded.

collectively exhaustive

When the list of attributes associated with a variable covers all possibilities. The response categories for a particular survey question should be collectively exhaustive (also referred to as just "exhaustive").

community-engaged research

A type of research that involves people in the communities being studied as active participants in the research process, rather than passive "subjects" or sources of information.

concept map

A visualization of how concepts relate to one another, which typically includes boxes that represent concepts and arrows that represent relationships (with the direction of the arrows indicating the presumed direction of causality).

concept mapping

Visualizing concepts and the relationships between them, generally by using boxes and arrows (with the direction of the arrows indicating the presumed direction of causality).

concepts

Mental images of a particular phenomenon that summarize its key aspects.

conceptualization

The stage of the research process at which researchers explicitly and clearly define the concepts they are using in their study.

confederates

Individuals who pretend to be research participants but are actually part of the research team running an experiment. Confederates interact with participants and otherwise behave in ways to establish the desired conditions for the experimental setting.

confidentiality

When researchers collect identifying information about research participants, but only they can link specific participants with that information.

confirmability

A criterion used to assess the rigor of qualitative research that focuses on the degree to which the results reported

are actually grounded in the data obtained from study participants. Confirmability is a dimension of the larger criterion of *trustworthiness*.

confirmation bias

A natural tendency to interpret data in ways that support, or "confirm," one's existing views, which can lead to flawed research findings that reflect the researcher's personal biases.

conflict of interest

Situations in which the financial or other personal benefits that researchers have received, or could receive, might compromise or bias their research by influencing their professional judgment and objectivity.

conflict theory

A major theoretical approach that is interested in questions of power. Scholars in this tradition examine who wins and who loses based on the way that society is organized.

confounder

A variable other than the presumed independent variable that may be influencing the dependent variable. (Also known as a *confounding variable* or *lurking variable*.)

consent form

A document given to research participants that provides clear and accurate information about the study, including any potential benefits and risks of the research and the procedures for storing personal data and withdrawing from the study. Participants are asked to sign the consent form or otherwise indicate they have read it and fully understand its contents.

content analysis

A materials-based research method that focuses on texts and their meanings. In content analysis, the "texts" being analyzed can be any recorded communications—from news articles or email messages, to speeches and dance performances, to television shows, advertisements, and movies.

content validity

A method used to assess the validity of a measure where a researcher evaluates whether the measure covers all the possible meanings, domains, and dimensions of a concept.

control group

A sample of research participants who do not receive the experimental stimulus but are included in a study to measure their scores on the dependent variable. Their outcomes can be compared to those of the experimental group to gauge the impact of the stimulus or treatment.

control variables

Variables of secondary importance that still need to be accounted for in a quantitative data analysis to avoid being fooled by spurious relationships between the independent and dependent variables.

convenience sampling

A sampling technique in which a researcher draws a sample from part of the population that is convenient to obtain—for instance, because potential interviewees are located near the researcher or otherwise are readily available.

convergent validity

A method used to assess the validity of a measure by comparing scores on that measure to those derived from an existing measure of the same or a similar concept. A strong correlation between the two measures is evidence that they are both measuring the same thing, and that the new measure is therefore valid in this way. (Compare to *discriminant validity*.)

correlation

When variables are related to one another, in the sense that changes in one variable are associated with changes in another variable. (Correlation is also referred to as *association*.) Note that observing that two variables are correlated is not by itself evidence that changes in one variable cause changes in the other (i.e., correlation does not necessarily mean causation).

correlation coefficient

A statistical measurement of the extent to which two ordinal-level or scale-level variables are linearly related to one another.

covert research

When a researcher does not disclose their identity as a researcher to the individuals or groups being studied.

credibility

A criterion used to assess the rigor of qualitative research that focuses on the degree to which the results are accurate and viewed as important and believable. Credibility is similar to validity in that it assesses how closely the study reflects the reality of the people being studied. It is a dimension of the larger criterion of *trustworthiness*.

critical theory

A paradigm of scientific knowledge focused on power, inequality, and social change. Critical theory attempts to analyze the oppositions, conflicts, and contradictions in contemporary society, critique the alienating and oppressive conditions of the status quo, and eliminate the causes of that alienation and oppression.

curriculum vitae

A document listing a scholar's credentials, research interests, teaching experience, and important publications and presentations. (Abbreviated as CV. The plural is also written as *curriculum vitae*.)

data

Facts, details, statistics, and other information used to understand some phenomenon of interest. Ideally, data is collected through a scientific process, but any information could be considered data for research purposes.

data analytics

Using scientific methods to interpret available information, identify trends and patterns, and map complex relationships.

data point

A specific statistic or other piece of information employed in research.

data science

An interdisciplinary field that uses computer coding and social scientific analysis methods to address the data-related problems facing corporations, governments, and other organizations.

dataset

All the data relating to a specific survey or study that is grouped together in a single file or set of files (referred to more colloquially as a *database*).

debriefing

A procedure for ensuring informed consent when research participants were intentionally deceived about the nature of a study or not provided with full information. In the debriefing held after the conclusion of the study, researchers explain its true purpose and why any deception was necessary.

deductive approach

An approach to empirical investigation in which researchers start with a social theory that they find noteworthy and then test its implications with data. (Also referred to with the terms *deduction* or *deductive analysis*.)

demography

A sociological field that focuses on the statistical study of populations.

dependability

A criterion for assessing the rigor of qualitative research that focuses on whether the researcher followed proper procedures in conducting the project. Dependability is a dimension of the larger criterion of *trustworthiness*.

dependent variable

A variable thought to be influenced or changed by another variable (the independent variable). Dependent variables are also referred to as *response variables*, *outcome variables*, and *outcome measures*.

descriptive field notes

Field notes that focus on straightforwardly describing a researcher's observations rather than analyzing them.

descriptive research

A type of research directed at making careful observations and generating detailed documentation about a phenomenon of interest.

descriptive statistics

Calculating statistics (e.g., counts, percentages, means) to summarize the main features of a sample or subsample. (Compare to *inferential statistics*.)

deviant cases

Particular cases (e.g., persons, organizations, observations) that contradict an observed pattern or existing theory, such as by having exceptionally high or low values on a noteworthy variable. (Also known as *outliers*.)

digital ethnography

A qualitative method of research that applies ethnographic observation techniques to online social spaces (e.g., social media platforms, video streams). Digital ethnographers examine people's virtual activities and interactions—sometimes in real time, and sometimes complemented by offline observations.

digital object identifier

A unique string of numbers and letters assigned to a scholarly paper, government report, or other document (abbreviated as *DOI* or *doi*). The nonprofit DOI Foundation introduced this international standard in 2000, and most academic journal articles are now assigned DOIs when they are published. Going to **https://doi.org/[DOI]** (with the article's DOI replacing **[DOI]** in the address bar) will take you to the official webpage for that source.

dimensions

Aspects of a concept that can vary.

direction of causality

A description of the presumed relationship between two variables that specifies whether changes in the first variable cause changes in the second, or vice versa.

direction of the relationship

Whether the overall relationship between two numerical variables is *positive* (as one goes up, the other goes up) or *negative* (as one goes up, the other goes down).

disclosure continuum

The range of possible disclosure that an ethnographer might provide about their role as a researcher in the field, which can be understood as a spectrum rather than discrete categories of "disclosure" and "nondisclosure." (See also *covert research*.)

discriminant validity

A method of assessing the validity of a measure by comparing scores on that measure with those derived from an existing measure that, logically speaking, should be entirely unrelated. The lack of any correlation between the two measures is evidence that they are measuring different concepts. (Compare to *convergent validity*.)

disinformation

Information that is deliberately misleading. (Compare to *misinformation*.)

disproportionate stratified sampling

An approach to stratified sampling in which the sizes of the subgroup samples do not match their relative sizes within the population (or sampling frame).

dissertation

An original research project that a PhD student typically must complete to receive their doctoral degree. PhD students who have successfully completed their coursework and any qualifying exams and are now working on their dissertation are called *PhD candidates* (and are colloquially described as *ABD*—"all but dissertation").

double-blinding

A procedure of keeping identities secret on both sides. In peer review, the reviewers are "blind" to the identities of the authors, and vice versa. In experiments, neither the researchers nor the study participants know who received the treatment until after the data collection is completed.

dyads

Pairs of people who are connected through social interaction.

ecological fallacy

A logical error of making claims about the nature of individuals based on data from the groups they belong to.

edited volume

A book whose chapters are scholarly papers written by various authors. One or more scholars will serve as editors of the volume, which typically focuses on a specific research-related theme or question.

emergent phenomena

Newly emerging topics or issues that researchers have not yet studied intensively.

empirical evidence

Data collected through a scientific process.

empirical generalizations

Patterns or regularities identified in the data that a researcher analyzes.

empirical papers

Papers that report the results of a quantitative or qualitative data analysis conducted by the author, oftentimes with original data collection as well.

empirical questions

Questions that have to do with our factual reality and that can be answered through research.

epistemology

Philosophical inquiry relating to the nature and scope of knowledge.

error

Inaccuracy in the measurement of key variables for a given population.

ethnographers

Researchers who immerse themselves in a particular setting, organization, or culture and conduct ethnographic observations (first-hand observations and descriptions of a social context), usually over an extended period of time. (Also known as *field researchers*.)

ethnographic observation

A qualitative method of studying a phenomenon within its social context by doing first-hand observations and providing detailed descriptions.

ethnography

A qualitative method of studying a phenomenon within its social context by doing first-hand observations and providing detailed descriptions. The word "ethnography" can also refer to studies that utilize ethnographic observation, including the books that ethnographers write based on such research.

evaluation research

A type of research that aims to assess whether social programs are effective in achieving their objectives.

exclusion criteria

Criteria that a researcher uses to decide whether to exclude a person (or other unit of analysis) from a sample.

experimental data

Information that researchers collect under specific conditions that they have some control over, or whose existence they can exploit.

experimental group

A sample of research participants who do receive the experimental stimulus or treatment. (Also called the *treatment group*.) Their scores on the dependent variable can be compared to those of the control group to gauge the impact of the stimulus or treatment.

experiments

A research method that ideally entails having control and experimental groups that are exactly the same except for the presence of a stimulus or treatment, allowing researchers to rigorously assess whether a change in the corresponding independent variable brings about a change in the dependent variable or variables of interest.

explanatory codes

Codes that go beyond merely categorizing data and attempt to make theoretical arguments. (Also called *analytic codes*.) During the coding process, codes normally shift from being more descriptive to being more explanatory.

explanatory research

A type of research that seeks explanations of observed behaviors, problems, or other phenomena. Explanatory research seeks answers to "why" and "how" questions.

exploratory research

A type of research that examines new areas of inquiry, with the goals of (1) scoping out the magnitude or extent of a particular phenomenon, problem, or behavior; (2) generating initial ideas or hunches about that phenomenon; or (3) testing the feasibility of undertaking a more extensive study regarding that phenomenon.

extended case method

An alternative strategy for qualitative research that does not follow the purely inductive approach of grounded theory. When using the extended case method (also known as the *extended case study approach*), researchers start with an existing theory and look for one or more cases that deviate from that theory, which they can use to explain why the theory falls short. Observations of the chosen case or cases are used to find and correct problems in the theory.

external validity

A study's ability to generalize any results obtained from its sample to its target population (or, more generally, to other people, organizations, contexts, or times).

extreme cases

Organizations or other sites of study that display characteristics or behaviors that are not the norm. These deviant cases (or outliers) stray from commonly seen patterns in instructive ways—for instance, by allowing researchers to see aspects of a phenomenon of interest that are rarer in conventional cases, or explain why existing theories do not adequately account for such an outlier (and therefore need to be refined).

face validity

A method of assessing the validity of a measure in which the researcher evaluates whether it is plausible and logical that a given variable (specifically, the operational definition of a concept) actually measures what it intends to measure.

fairness

A criterion used to assess the rigor of qualitative research that focuses on the extent to which researchers have considered all relevant viewpoints on a particular topic. Fairness is a dimension of the larger criterion of *authenticity*.

falsifiable

The possibility that empirical observations can show a hypothesis generated from a theory to be untrue. The British philosopher Karl Popper argued that *falsifiability* is a foundation of the scientific method: for theories to be valid, they must be falsifiable.

feedback loops

Situations in which a change in concept A leads to a change in concept B—which, in turn, loops back to change concept A.

field experiments

Experiments conducted in a real-world setting rather than in a laboratory.

field jottings

Descriptive notes that researchers write—in a notepad or through more discrete means—while they are observing in the field or during an interview. Their field jottings supply the raw material that they later use to draft more formal field notes.

field note

A memo that researchers write to describe and summarize their observations (and sometimes also the content of their interviews) during a particular period of time. Field notes (also written as *fieldnotes*) often contain the researchers' personal reactions and preliminary analysis as well. To ensure adequate recall, researchers try to write field notes right after their last session of data collection, drawing on the field jottings they wrote while observing.

field site

A setting, organization, or other social context that a researcher chooses to study in order to understand a phenomenon. (Also referred to as just a *site*.)

focus groups

Interviews conducted with a group of respondents at the same time. During a focus group session, one or more moderators will typically ask the group questions about a particular political issue, product, or other topic.

focused coding

A stage of the coding process that typically follows open coding. During focused coding, researchers finalize their code lists—collapsing or narrowing their codes and setting firm definitions for each one—and then recode their qualitative data to ensure consistency.

folk theories

Explanations for phenomena that the groups being studied believe in. Folk theories can be useful as sources of insight and insider knowledge, but they also may reflect the group's prejudices or blind spots.

foundations

Private organizations that manage large amounts of assets on behalf of a particular public purpose.

generalist journals

Academic journals that publish any noteworthy paper within a broad field.

generalizability

The extent to which a study's results can reasonably tell us something about the larger population from which its sample was drawn.

generalizable

When a study's results can reasonably tell us something about the larger population from which its sample was drawn.

gray literature

Research and information produced by nonacademics, including researchers working for government agencies, advocacy organizations, polling outfits, and think tanks.

grounded theory

A purist approach to inductive investigation in which researchers start from a clean slate, letting data guide them to new theories rather than beginning with a set of existing theories to build on or test. By keeping an open mind while immersing themselves in real-world settings, researchers following the grounded theory approach should be able to generate novel theories that are not constrained by any preconceptions or well-established views.

history threat

The possibility that an observed change in the dependent variable is caused by extraneous or past events rather than by the experimental treatment. The history threat (also called the *history effect*) is one kind of threat to the internal validity of an experiment.

human subject

A living individual about whom a social scientist (whether a professional or student) either (1) obtains information through intervention or interaction with the individual, and uses, studies, or analyzes the information; or (2) obtains, uses, studies, analyzes, or generates identifiable private information.

hypothesis

Scientific conjectures—educated guesses—about how the various concepts being studied are related, which researchers develop based on logic or the findings of past research.

ideal types

The types present within a typology, which are not meant to describe any one particular case but rather to highlight commonly shared characteristics across cases. The German sociologist Max Weber popularized the use of ideal types.

idiographic explanations

When a researcher explains a single situation or event in idiosyncratic detail by listing all its potential causes.

impact factor

A rough numerical measure of a journal's influence within the scientific community, which is calculated based on how many citations its articles received, on average, a year or two after publication.

implementation research

A type of research that attempts to make better use of scientific knowledge in real-world settings, paying close attention to the roles that various groups (e.g., governments, nonprofit organizations, healthcare providers, citizens) play in the successful implementation of that knowledge.

in vivo codes

Codes that emerge from the terms or phrases that research participants themselves use.

in-depth interviewing

A qualitative method of research that involves conducting semi-structured or unstructured interviews, with a focus on asking individuals open-ended questions to elicit detailed information. (Also called *qualitative interviewing*.)

in-depth interviews

Semi-structured (and, more rarely) unstructured interviews that focus on generating rich qualitative detail about a topic. (Also called *qualitative interviews*.) During in-depth interviews, researchers prioritize open-ended questions (which give respondents more flexibility to discuss what they think is important) and probes (predetermined or improvised follow-up questions).

inclusion criteria

Criteria that a researcher uses to decide whether to include a person (or other unit of analysis) within a sample.

independent scholars

Researchers without any affiliations to a university, think tank, or other institution.

independent variable

A variable that a researcher believes explains changes in another variable. Specifically, changes in the independent variable are thought to cause changes in the other variable (the dependent variable). Independent variables are also known as *explanatory variables*. In experiments, the independent variable that the researcher manipulates is called the *experimental stimulus* or *treatment*.

index

A type of measure that contains multiple indicators designed to assess a more general concept. Although the terms "index" and "scale" are sometimes used interchangeably, indexes typically have items that are different conceptually from each other, rather than collectively measuring intensity across a single dimension.

indicators

Another term for *variables*—operational definitions of concepts—that is frequently used in quantitative research, especially for variables that are routinely collected by government agencies and other organizations.

inductive approach

An approach to empirical investigation in which researchers start with a set of observations and use the empirical evidence they gather to create a more general set of propositions about how the world operates. (Also referred to as *induction* or *inductive analysis*.)

inference

The process of testing hypotheses based on empirical evidence.

inferences

Conclusions based on empirical evidence.

informal observation

When people make observations without any systematic procedure for conducting their observations or assessing the accuracy of what they observed.

informed consent

In research ethics, the principle that research participants should be included in a study only after they have been provided easily understandable and accurate information about it, and only after they have willingly given their consent to participate.

institutional review board

Board of experts established at universities or other research organization to review research proposals involving human subjects and ensure that researchers follow ethical principles. Review by institutional review boards (IRBs) is a requirement for any federally funded research in the United States; in the United Kingdom, similar regulatory bodies are known as *ethics committees*.

instrumentation threat

The possibility that the difference between an experiment's pretest and posttest scores could be due to changes in the administered test. For example, the pretest might have a lower degree of difficulty than the posttest, thus accounting for increased scores. The instrumentation threat (also called the *instrumentation effect*) is one kind of threat to the internal validity of an experiment.

inter-rater reliability

A method of assessing the reliability of a measure by examining the degree to which two or more observers (*raters*) agree on the measurement of one or more cases (i.e., on the values assigned to those cases).

intergovernmental organizations

Organizations that involve two or more nations working to further a common interest or goal.

internal consistency

The degree to which participants' answers to items within a multiple-item measure are consistent. Specifically, the answers for each item in an index or scale should be correlated with each other, as they all are supposed to measure aspects of the same overall concept.

internal validity

A study's ability to determine if changes in the independent variable truly cause changes in the dependent variable.

interpretive approach

An approach to research that attempts to make sense of what people say or do, with a focus on the social significance of those words and actions. (Sometimes referred to as the *interpretivist approach*.)

interval level

A level of measurement in which the variable's attributes are separated by equal and meaningful distances. Interval-level variables are included in the broader category of *scale-level* variables. Unlike ratio-level variables (also in that

category), interval-level variables do not have a *true zero point*, where a value of zero refers to the complete absence of the phenomenon being measured.

interview guide

A list of topics or questions that the interviewer can refer to during the course of an interview.

interviews

A method of data collection that involves two or more people exchanging information through a series of questions and answers.

iterative sampling

A sampling strategy in which researchers move back and forth between the process of sampling and preliminary analysis of their data. (Also known as *recursive sampling*.)

key informants

Research participants who are well-connected within the setting that the researcher is studying. Such an insider can provide a framework for your observations, help "translate" what you observe, give you important insight into a group's culture, and assist you in recruiting additional participants to observe or interview.

keywords

When searching the scholarly literature on a topic, keywords refer to relevant terms used in a search query. When writing a journal article, authors choose a number of terms to be explicitly listed as the article's "keywords," which describe the main concepts and topics the paper covers.

laboratory experiment

An experiment done in an office or other setting where researchers can manipulate conditions to emulate situations in the real world.

laws

Consistently observed patterns relating to specific phenomena.

leading question

When the wording of a question encourages a person to answer in a potentially biased way.

level of measurement

A classification system that categorizes variables by how their attributes are related to one another. In statistical analysis, researchers focus on three key levels of measurement: *nominal*, *ordinal*, and *scale*. Different analysis techniques are appropriate for each level of measurement.

levels of analysis

The different levels of aggregation at which social scientists can study phenomena, which can range from the *macro* (societies) to the *micro* (individuals).

life history

A portion of an in-depth interview focused on formative events in the respondent's life that may have influenced their orientation toward a particular issue or their experience of a particular phenomenon today.

Likert scales

Ordinal-level scales used to measure the intensity of people's agreement with a particular statement. A common format for a Likert scale (pronounced LICK-ert) has response categories ranging from "Strongly Disagree" to "Strongly Agree," corresponding to numerical values of 1 to 5 (including a neutral value) or 1 to 4 (without a neutral value).

linking concept

A concept that describes a pathway by which concept A (or the independent variable) affects concept B (or the dependent variable). (Also referred to as a *causal mechanism*, *mediating concept*, or mediator.)

literature

The existing scientific studies that relate to a particular phenomenon.

literature review

A summary, analysis, and synthesis of the most significant published research on a scholarly topic.

literature review matrix

A grid, table, or outline that lists all the sources to be included in a literature review, breaking them down by their research questions, methods, samples, key findings, and other characteristics. (Also called a *summary table*.)

macro level

The highest level of analysis, which in the social sciences often involves the study of communities, societies, or countries.

manual coding

Conducting an analysis of qualitative data without the use of specialized qualitative data analysis (QDA) programs.

market research

Research to help guide companies make decisions about increasing their sales or improving their performance.

maturation threat

The possibility that a change in the dependent variable observed within an experiment could be caused by natural changes in research participants rather than by the experimental treatment. The maturation threat (also called the *maturation effect*) is one kind of threat to the internal validity of an experiment.

measurement error

The difference between the measured values of a variable and the true but unobserved values of that variable. The two kinds of measurement error are *random error* and *systematic error*.

measures

Another term for *variables* (operational definitions of concepts).

mediating variables

Variables that describe a pathway by which an independent variable affects a dependent variable. (Also referred to as *causal mechanisms*, *linking concepts*, or *mediators*.)

mediation

When a concept stands between and links two other concepts in a causal relationship. A *mediating concept* (or *linking concept*) is the pathway by which one concept affects another.

mediator

A concept that describes a pathway by which concept A (or the independent variable) affects concept B (or the dependent variable). (Also referred to as a *causal mechanism*, *linking concept*, or *mediating concept*.)

meso level

The middle level of analysis, which in the social sciences typically involves the study of organizations or other kinds of groups.

meta-analysis

A common type of systematic review that uses statistical procedures to synthesize the results of many quantitative studies. (Also called a *meta-synthesis*.)

micro level

The lowest level of analysis, which in the social sciences typically involves the study of individuals.

misinformation

Information that is inaccurate or misleading. (Compare to *disinformation*.)

mixed-methods research design

A research design that uses qualitative and quantitative techniques jointly within a single study. (Also referred to as *mixed methods* or a *mixed-methods approach*.)

moderating variables

Variables that alter the relationship between dependent and independent variables.

moderation

When a concept (or variable) influences the relationship between two other concepts (or variables). (Also referred to as *interaction* or an *interaction effect*.) Specifically, the presence of this *moderating concept* (also called a *conditioning concept*) weakens or strengthens (or otherwise affects) the relationship between two concepts.

moderator

The researcher tasked with managing the conversation in a focus group. (Also called a *facilitator*.) The moderator

poses questions and interjects as needed, with the goal of getting group members to share their thoughts and interact with one another during the session.

monographs

Scholarly books that focus on a particular topic.

mosaic plagiarism

A form of plagiarism in which the author cites the source that they drew ideas or information from, but nevertheless uses a substantial amount of the source's exact wording without explicitly quoting the copied passages.

multidimensional concepts

Overarching concepts formed from multiple underlying concepts.

mutually exclusive

When the different attributes defined for a variable (or response options provided for a survey question) do not overlap with one another. All variables used in quantitative data analysis must have mutually exclusive attributes.

natural sciences

Scientific fields that focus on naturally occurring objects or phenomena, such as light, objects, matter, earth, stars, or the human body. (Also known as the *hard sciences*.)

needs assessment

A systematic effort to determine what assistance may be needed by the individual clients or communities that a program serves.

negative relationship

A type of relationship between two numerical variables in which the value of one variable goes down as the value of the other variable goes up, and vice versa. (Also called an *inverse relationship*.)

neologisms

Newly coined words or expressions, which a researcher can devise to capture a new or enhanced way of understanding a particular phenomenon.

network-based referrals

Recruiting potential interviewees by asking people in a relevant social network to provide referrals. This approach can entail *snowball sampling* (asking your interviewees to suggest other potential interviewees) and *organizational referrals* (getting people who work at a particular organization to refer you to members of the population that their organization serves).

no harm to participants

In research ethics, the principle that participants should not suffer physical or psychological harm.

nominal level

The lowest level of measurement, where a variable's attributes are different from one another, but there is no mathematical order to the attributes. City of residence and gender identity are examples of nominal-level variables.

nomothetic explanations

When a researcher attempts to explain a class of situations or events rather than a specific situation or event.

nongovernmental organizations

Nonprofit organizations that operate outside of government control and typically conduct operations on behalf of social or political causes across multiple countries. (Abbreviated as NGOs.)

nonprobability sampling

A type of sampling in which the researchers do not know the likelihood that a person (or other unit of analysis) in the population will be selected for membership in the sample. Nonprobability sampling is common in qualitative research.

nonprofit organizations

Organizations like charities, churches, labor unions, and political organizations that are not government-owned and operate within the private sector, yet—unlike for-profit firms—commit any profits they make to the organization's social purpose.

nonrespondents

Individuals who decline to answer questions for a survey or in-depth interview or who cannot be reached by researchers.

nonresponse bias

Bias introduced into a study when respondents and nonrespondents differ in important ways, which means that the relevant characteristics observed in the sample differ from those in the target population.

nonspuriousness

When no alternative explanation exists for the observed result (the dependent variable) other than the action of the independent variable. Nonspuriousness is a necessary condition for establishing a causal relationship between two variables.

nonverbal cues

A respondent's facial expressions, gestures, and other types of body language, which can give hints about their mood or their underlying feelings and thoughts about a topic they are discussing.

normative questions

Questions that concern what norms or standards society should have, and whose answers therefore depend on people's moral opinions. Research can inform, but not answer, normative questions.

Nuremberg Code

Ten principles for ethical research put forward by the Nuremberg war crimes tribunal in 1947 to guide medical professionals and scientists who conduct research on human subjects.

observational data

Information that scientists gather in the real world from surveys, in-depth interviews, ethnographic observation, and other methods that do not use an experimental design.

observations

First-hand experiences of some phenomenon.

ontology

Philosophical inquiry into the nature of reality.

op-eds

Commentary pieces published by news outlets. (Also known as *guest essays*.) "Op-ed" stands for "opposite the editorial page," referring to where such guest essays were traditionally placed in a newspaper.

open coding

An inductive process of reading through qualitative data, categorizing the material by relevant themes, and developing theoretical arguments based on the observed patterns or relationships.

open-ended questions

Interview questions that give respondents more leeway to answer as they wish. In in-depth interviews, open-ended questions encourage interviewees to answer at length. In surveys, open-ended questions allow respondents to respond using their own words rather than choosing from response categories.

operational definitions

A definition or procedure for how researchers actually measure an abstract concept when they are collecting data.

operationalization

The stage of the research process at which the researcher specifies explicitly and clearly how a concept will be measured.

operationalize

When a researcher specifies how a concept will be measured.

ordinal level

A level of measurement in which a variable's attributes can be numerically ranked from low to high values (or vice versa) using some kind of meaningful comparison. Unlike for scale-level variables, the distances between values of an ordinal-level variable do not correspond to actual differences in intensity or magnitude, and the values themselves only refer to the relative ranking of the variable's attributes.

outcomes assessment

An analysis conducted to determine whether a program has achieved its intended outcomes. Typically, an outcomes assessment measures the various resources the program took advantage of (*inputs*) and its tangible results (*outputs*).

outliers

Particular cases (e.g., persons, organizations, observations) that contradict an observed pattern or existing theory, such as by having exceptionally high or low values on a noteworthy variable. (Also known as *deviant cases*.)

overgeneralization

Assuming that broad patterns exist based on very limited observations.

oversampled

When a subgroup represents a greater share of a sample than the same subgroup represents in the larger population.

paper

A write-up of a study that a researcher has conducted.

paradigms

Ways of viewing the world and understanding the human experience. Paradigms frame what we know, what we can know, and how we can know it.

parallel literatures

Bodies of research that are not directly on your study's topic but overlap in some key way, which means they may have helpful concepts, theories, or empirical findings that you can apply to your own work.

parameters

Means, medians, standard deviations, and other numerical values that summarize the characteristics of a population. When researchers cannot observe the actual population parameters, they generate statistics from a sample of that population to estimate those parameters.

parsimony

How economical a theory is, as measured by how much of a particular phenomenon it can explain, with how few variables.

participant observation

A type of ethnographic observation where researchers get involved in the activities or organizations they are studying, taking on more or less formal roles as event participants or members of groups.

participation continuum

The range of possible involvement that a researcher might have in the activities or organizations they are observing in the field, which can be understood as a spectrum rather than discrete categories of "participation" and "nonparticipation."

participatory action research

A type of research that seeks to improve the conditions of a particular community. (Also referred to as just "action research.") Rather than merely measuring and explaining a social problem facing the community, action research seeks to identify and implement specific solutions to that problem.

peer review

A formal process in which other researchers review a scholarly work to ensure that it meets the standards and expectations of their field.

periodicity

When a sampling frame exhibits a pattern that occurs at regular intervals, such as weekends occurring in a sampling frame of dates. Under such conditions, systematic sampling might inadvertently introduce bias into a sample.

phenomena

Things, events, and processes that can be observed and that raise interesting questions for scientists about how the world works.

pilot study

Research that explores initial ideas or tests out data collection or analysis procedures as part of a more comprehensive project.

pilot testing

Any preliminary vetting of a survey questionnaire, interview guide, or other research instrument.

plagiarism

The presentation of someone else's work as your own. To avoid plagiarism, scholars must cite all sources that they took ideas or information from (both within the text and in the reference list at the end of their paper), put quotation marks around any text that was used word-for-word, and avoid paraphrases that are too close in wording to the text of the original source (even when it is cited).

policy briefs

Reports written by experts that summarize a particular issue and the recommended policy approaches for dealing with it.

policy intervention

Action by policymakers to launch or expand particular programs or enact or alter particular laws.

policymakers

Top government, corporate, or other officials who can create laws, regulations, rules, and other sorts of policies that affect people's behavior.

population

The larger group that a sample of people (or other units of analysis) are drawn from, and that the sample is intended to represent.

population of interest

The larger group (of people, organizations, objects, etc.) that a researcher is interested in learning about and that their research question applies to. (Also referred to as a *target population*.)

positionality

How a researcher's personal identities and perspectives shape how they interpret reality, including what they observe over the course of their research and what conclusions they draw.

positive relationship

A type of relationship between two numerical variables in which the value of one variable goes up as the value of the other variable goes up, and vice versa.

positivism

A paradigm of scientific knowledge that prioritizes principles of objectivity, knowability, and deductive logic.

postmodernism

A paradigm of scientific knowledge that directly challenges the idea of objective knowledge as well as the various ways of knowing that science has traditionally used.

postpositivism

A paradigm of scientific knowledge that, like positivism, seeks to reject false beliefs through empirical observation, but also holds that it is impossible to verify the truth in any exhaustive fashion.

postsurvey weighting

Performing statistical weighting to adjust the sample after it has been collected and make its statistics better reflect the actual parameters of the target population.

posttest

A measurement of the dependent variable taken after the stimulus or treatment has been introduced in an experiment.

pre-experimental designs

Experimental designs that do not have the conventional features intended to ensure *internal validity* (the study's ability to evaluate whether a causal relationship exists between two variables). These features would include having study participants randomly assigned to experimental and control groups, providing the stimulus or treatment only to the experimental group, and conducting pretests and posttests (measurements of the dependent variable).

predictive validity

A method of assessing the reliability of a measure by determining if it predicts future phenomena that it should be able to predict. For example, a standardized test that successfully predicts a student's grades in college would arguably have predictive validity as a measure of academic ability.

pretest

A measurement of the dependent variable taken at the outset of an experiment before the stimulus or treatment has been introduced.

primary sources

Original works describing first-hand observations, often written by individuals who were present at a noteworthy event or had relevant experiences.

principal investigators

Researchers who have a central role in conducting a study. (Abbreviated as PIs.)

private sector

The portion of an economy (including the labor market) that is not controlled by the government—for instance, the activities of for-profit firms and nonprofit organizations.

probability proportionate to size

In cluster sampling, a procedure to adjust the sample selection process so that each element across the differently sized clusters has an equal chance of being selected. (Abbreviated as PPS.)

probability sampling

A type of sampling in which the researchers know the likelihood that a person (or other unit of analysis) in the population will be selected for membership in the sample. (Also called *random sampling*.) Probability sampling is widely used in quantitative studies.

probes

Follow-up questions that a researcher asks during an in-depth interview, which may be part of the interview guide or improvised.

professional associations

Membership organizations that represent individuals who share employment or interests in a specific industry or field, such as an academic discipline like sociology.

professional codes of conduct

Sets of formal ethical rules and principles developed and codified by professional organizations, such as the American Sociological Association.

program

A set of interventions that individuals or groups receive, often as part of an organization or government's strategy to deal with a social problem.

proportionate stratified sampling

An approach to stratified sampling where the sizes of the subgroup samples match their relative sizes in the population (or sampling frame).

protection of identities

In research ethics, the principle that researchers should safeguard the privacy of research participants through approaches such as ensuring confidentiality or anonymity.

public health

An interdisciplinary science focused on studying the determinants of health—including social determinants—and using that knowledge to craft policies and interventions that prevent disease and promote health within communities and societies.

public sector

The portion of an economy (including the labor market) that is controlled by the government.

public sociology

The application of sociological theories and research to matters of public interest.

purposive sampling

A nonprobability sampling approach where the selection of cases is guided by the researcher's theory about what concepts and processes matter. (Also called *theoretical sampling*.)

qualitative methods

Types of research that generate or analyze data involving words, pictures, and other symbols beyond numbers. Two of the most common qualitative methods in sociology include ethnographic observation and in-depth interviewing.

quantitative methods

Types of research that generate or analyze data that can be represented by and condensed into numbers. Survey research is probably the most common quantitative method in sociology, but methods such as content analysis can also be conducted in a way that yields quantitative data.

quasi-experiments

Studies that approximate a true experimental design to improve their internal validity beyond what they could do with merely an analysis of observational data. However, quasi-experimental designs lack random assignment to experimental and control groups, which means they are vulnerable to related threats to internal validity.

queries

Search terms used in a database to find sources.

quota sampling

A nonprobability sampling technique where a researcher starts by identifying categories that are important to the study. The researcher then selects a sufficient number of cases that fall into these key categories, sometimes in proportion to the relative size of the corresponding subgroups within the population.

R1 schools

Large research universities that engage in "very high research activity," according to the Carnegie Classification of

Institutions of Higher Education, which means that they have substantial doctoral programs and numerous faculty and students focused primarily on research.

random error

A type of measurement error—present in every measurement—where unpredictable differences occur between the measured values of a variable and the true but unobserved values. (Compare to *systematic error*.)

random sampling

A sampling process that ensures that cases from the population are picked at random.

random selection

A sampling process that ensures that cases from the population are picked at random.

rapport

The sense of connection a researcher establishes with a research participant, which may encourage the participant to speak in greater detail about their thoughts, feelings, and experiences or to act more naturally while being observed.

ratio level

The highest level of measurement, where the variable's attributes are separated by equal and meaningful distances (like for interval-level variables) and the variable has a true zero point (unlike for interval-level variables). For example, the number of siblings a person has would be a ratio-level variable, given that a value of zero refers to the complete absence of the phenomenon being measured (i.e., not having any siblings). Ratio-level and interval-level variables are both included in the broader category of scale-level variables.

reactivity

The degree to which the presence of a researcher or their interactions with participants alter the behaviors of those being observed, leading to outcomes different from what would have happened without the researcher present.

refereed journal

An academic journal whose papers are vetted through peer review. (Also known as a *peer-reviewed journal*.)

reference group

The social group to which people compare themselves when judging their own social status, such as when they assess their income relative to others.

reflexivity

A self-reflective process that researchers engage in to understand how their own identity, beliefs, dispositions, actions, and practices may have influenced their research, especially the results they found.

regression to the mean

The statistical tendency of a group's overall performance on a measure during a posttest to move toward the mean of that measure rather than in the hypothesized direction. For example, if participants score very low on a pretest, they will have a tendency to score higher on the posttest (closer to the mean). Regression to the mean (also called

regression, though note that this usage is distinct from "regression" in the sense of *regression analysis*) is one kind of threat to the internal validity of an experiment.

reification

The assumption that an abstract concept exists in some concrete, tangible way.

reliability

The consistency of a measure. A measure is said to be reliable if it gives the same result upon repeated applications to the same phenomenon.

representative sample

A sample whose characteristics are similar to the population from which it was drawn, which means that findings from that sample can be reasonably generalized to the population.

research design

The planning process for a scientific study, which typically involves a thorough review of the relevant literature, the formulation of a focused research question, and a detailed proposal for the methodological approach that will be used to answer that question.

research instrument

Particular tools that are used in research to measure concepts, such as a survey questionnaire or interview guide.

research methods

Organized and logical ways of learning about and knowing the world around us.

research participants

The people or communities being studied by a researcher. (Also called *study participants* or just "participants.")

research problem

A gap in the literature on a topic that scholars who work in that area aim to address with their research.

research proposal

The end product of the process of research design for a proposed scientific study, which usually takes the form of a paper that includes a thorough review of the relevant literature, the formulation of a focused research question, and a detailed plan for the methodological approach that will be used to answer that question.

research question

The question a researcher hopes to answer by collecting and analyzing data for an empirical study.

respondents

Individuals who answer questions for a survey or in-depth interview.

response categories

The listed options that a respondent can choose when answering a particular question (*item*) on a survey

instrument. (Also called *response options*.) When constructing a survey questionnaire, each question's response categories should exhaust all the possible attributes for the associated variable.

response rate

A percentage determined by dividing the number of completed survey questionnaires by the number originally distributed, or the number of individuals successfully interviewed by the number contacted for an interview.

reverse causality

A situation in which researchers believe that a change in concept A (or the independent variable) causes a change in concept B (or the dependent variable), but the opposite is actually the case.

review articles

Long-form literature reviews that scholars write to synthesize and comment on the existing research on a particular topic of scientific interest. Some journals specialize in publishing review articles, such as the *Annual Review of Sociology*.

sample

The subset of the larger population that the researcher has collected data from, and that represent the target population within the study.

sample for range

A sampling strategy where the researcher seeks to find a sufficient number of cases that reflect the range of variation across one or more key variables.

sampled population

All the people (or other units of analysis) whom researchers seek to recruit from the population of interest.

sampling

The process of selecting cases or observations that will be analyzed for research purposes.

sampling bias

A type of bias that occurs when the elements selected for inclusion in a study do not represent the larger population from which they were drawn.

sampling error

The difference between the statistics obtained from a sample and the actual parameters of a population. Probability sampling allows for the calculation of the sampling error (also called *random sampling error*) that is expected given the size of the sample being used.

sampling frame

A list of members of a population that is available to researchers, which they use to select cases for their sample. Ideally, the sampling frame includes every single member of that population.

sampling unit

Each case in the researcher's sample—for example, each individual interviewed in an interview-based study.

sandwich method

An approach to writing up qualitative data that starts with a theoretical point (the top bun), inserts qualitative data (the meat), and then reiterates, expands upon, or notes exceptions to that theoretical point (the bottom bun).

saturation

Reaching a point at which collecting more data does not generate further insights into the phenomenon of interest.

scale

A type of measure that contains multiple indicators designed to assess a more general concept. Although the terms "index" and "scale" are sometimes used interchangeably, scales are different in that they have some kind of intensity structure. For instance, as a respondent moves from lower to higher values on the Bogardus Social Distance Scale, their level of tolerance for members of a particular group intensifies.

scale level

A level of measurement where the variable's attributes are ordered (like for ordinal-level variables) and the distance between those attributes is meaningful (unlike for ordinal-level variables). The scale level of measurement combines both the *interval* and *ratio* levels of measurement; for practical purposes, the same quantitative analysis techniques can be used for variables at either of these levels.

science

A discipline of learning that uses formal, systematic procedures to observe and measure phenomena.

scientific knowledge

A systematic and organized body of knowledge that is acquired using the scientific method.

scientific method

A standardized set of techniques for acquiring and vetting knowledge, such as procedures for making observations or interpreting results.

scope conditions

The conditions under which a relevant theory derived from a study's empirical research can and cannot reasonably be applied. (Also called *boundary conditions*.)

secondary data

Data that other researchers have previously collected. Frequently, government agencies and other organizations allow access to survey and census datasets for independent analysis, and independent researchers may provide their own data for others to analyze. Qualitative secondary data also exists, but it is much rarer in part because of confidentiality restrictions placed on such data.

secondary data analysis

Analysis of data that has previously been collected by other researchers.

secondary sources

Publications that interpret, discuss, and summarize primary sources.

selection criteria

The criteria that a researcher uses for including or excluding potential research participants from their sample. (See also *inclusion criteria* and *exclusion criteria*.)

selection interval

In systematic sampling, the distance between the elements of the sampling frame that researchers select for inclusion in their sample. The selection interval (k) is calculated by dividing the total number of population elements by the desired sample size.

selection threat

The possibility that differences between the control and experimental groups that existed prior to the administration of the stimulus or treatment could account for differences between those groups at the posttest. The selection threat (also called the *selection effect*) is one kind of threat to the internal validity of an experiment.

selective coding

A type of coding that involves identifying a central code or a core variable and systematically relating it to all other codes.

selective observation

When people see only those patterns that they want to see, or when they assume that only the patterns they have directly experienced actually exist.

self-selection bias

Bias that occurs when certain types of people are more likely to volunteer for (or be selected into) a sample. For example, people with strong opinions on an issue may be more likely to participate in a study.

semi-structured interviews

Interviews in which researchers have a list of questions prepared in advance, but they may deviate from that list whenever they wish, rephrasing or reordering their predetermined questions or improvising new ones on the spot. Most in-depth interviews are semi-structured.

signposting

Signaling the organization and structure of a paper (or presentation) to its readers (or audience) by stating and reiterating its key points or arguments.

simple random sampling

A sampling technique where the researcher gives all members of a population (more accurately, of a sampling frame) an equal probability of being selected.

single-blinded

A peer-review process in which the reviewers know who the author or authors are. For some peer reviews, such

as those used for academic books, the process is single-blinded for practical reasons (such as the length of the reviewed work making it impractical to mask the identity of the author).

snowball sampling

A sampling technique where researchers ask study participants they have already recruited to help identify additional participants.

social constructionism

A paradigm of scientific knowledge that sees "truth" as a varying, socially constructed, and ever-changing notion, and "reality" as created through people's interactions and interpretations of those interactions.

social desirability bias

Bias that occurs when participants in a research study answer or act in particular ways to present themselves to the researcher in a more positive light.

social determinants of health

The social and environmental conditions and patterns of social interactions that influence the well-being of individuals.

social network analysis

A method for examining the ties linking individuals, groups, or other units and the dynamics and consequences of these structures. (Abbreviated as SNA.)

social policy

A type of public (i.e., government) policy that has to do with human well-being, such as healthcare policies, education policies, and labor regulations.

social sciences

Scientific fields that focus on studying people or collections of people (such as groups, firms, societies, or economies) and their individual or collective behaviors.

sociology

The study of social life and social groups. Sociologists study how individuals shape (and are shaped by) social groups, which may be as small as individual families or couples, or as large as whole nations.

specialist journals

Academic journals dedicated to a subfield within the discipline, such as the sociology of education or organizational sociology.

spuriousness

A situation in which a relationship between two concepts seems to exist but, in reality, they are linked by a third concept, a *confounder* (also known as a *confounding variable* or *lurking variable*). Relationships where this condition holds are referred to as *spurious relationships*.

stakeholders

The various individuals or groups with a stake in the outcome of an endeavor. For a research project, the stakeholders would include the individuals or communities being studied, as well as scholars and policymakers who work on similar issues.

statistics

Means, medians, standard deviations, and other numerical values that describe a study's sample or a subset of it. Statistics are usually meant to be estimates of the corresponding *parameters* in the target population.

stimulus

The manipulation of an independent variable in an experiment, usually a specific intervention or factor whose impact on the dependent variable is what the researcher wishes to measure. (Also called the *treatment*.)

storylining

Taking theorized categories and relationships (as depicted in a concept map, for instance) and using them to tell a causal story about the observed phenomenon.

strata

In stratified sampling, the mutually exclusive subgroups that a researcher divides the sample into to ensure that the sample draws from each subgroup.

stratified sampling

A sampling technique where researchers divide the study population into two or more mutually exclusive subgroups (known as *strata*) and then draw a sample from each subgroup. Stratified sampling is used to ensure that the sample adequately represents the identified subgroups.

structural functionalism

A major theoretical approach that focuses on the interrelations between various parts of society and how each part works with the others to make society function in the way that it does—much like parts of the body work together to help an organism to thrive.

structured interviews

Interviews in which the researchers ask the exact same questions in a precise ordering that they planned out ahead of time. Survey interviews usually take the form of structured interviews.

supplemental questionnaire

A survey form that a researcher asks a respondent to fill out before or after an in-depth interview. A supplemental questionnaire can cover basic demographic information that the researcher does not want to go over during the interview itself, or it can ask about sensitive or hard-to-remember details that a respondent may be better able to answer outside the interview.

suppressor variable

A type of confounder that influences the dependent variable in such a manner that not accounting for it will

lead a researcher to mischaracterize the relationship between the independent variable and dependent variable as positive when it is actually negative, or vice versa.

survey interviews

Structured interviews in which the researcher poses the same set of questions, typically in a written format, to a sample of individuals.

surveys

A quantitative method of research (formally called *survey research*) that involves posing the same set of predetermined questions, typically in a written format, to a sample of individuals.

symbolic interactionism

A major theoretical approach that studies how meaning is created and negotiated through social interactions.

symbols

The representations that people use to communicate their thoughts and feelings.

systematic error

A type of measurement error that occurs when a measure consistently produces incorrect data. Systematic error typically occurs in one direction, in that the measured values are consistently too high or too low compared to their true values. (Compare to *random error*.)

systematic reviews

A synthesis of past research usually focused on a narrow empirical question. More common in medical and policy fields, systematic reviews typically attempt to make precise conclusions about the effect of a specific intervention. They usually draw on the existing quantitative literature, though some cover qualitative work.

systematic sampling

A sampling technique where the researcher selects elements from a sampling frame in specified intervals—for instance, every kth element on the list (where the selection interval k is calculated by dividing the total number of population elements by the desired sample size). To allow for an equal chance that every element could be selected, it is also important that the starting point be randomly chosen from within the first k elements on the list.

target population

The larger group (of people, organizations, objects, etc.) that a researcher is interested in learning about, and that their research question applies to. (Also referred to as a *population of interest*.)

tertiary sources

Publications that summarize secondary sources.

test-retest reliability

A method of assessing the reliability of a measure by collecting data from a sample and then retesting the same sample after a period of time. If the measure is reliable, the two measurements should be consistent.

testing threat

The possibility that a change in the dependent variable observed within an experiment could be caused by the participants' pretest responses influencing their posttest responses. The testing threat (also called the *testing effect*) is one kind of threat to the internal validity of an experiment.

The Belmont Report

A 1979 report produced by a U.S. federal agency, the National Commission for the Protection of Human Subjects in Biomedical and Behavioral Research, which laid out three overarching ethical principles to guide all federally funded research on human subjects: *respect for persons*, *beneficence*, and *justice*.

theoretical models

Theories that provide a simplified understanding of some process. Theoretical models (also just called "models") reduce a complex phenomenon into its most important parts.

theoretical papers

Papers that focus on elaborating a conceptual model or framework for understanding a problem rather than discussing data the author has collected or analyzed. (Also referred to as *theory papers*.)

theories

Systematic explanations of a natural or social behavior, event, or other phenomenon.

theory

A systematic explanation of a phenomenon.

thick description

A detailed description of the unfolding of a scene observed first-hand, with particular attention to the subjective and cultural meanings of any behaviors and other aspects of the larger social context.

think tanks

Private or government-run organizations that conduct research to help governments, corporations, social movements, or other groups make decisions about the best policies to pursue or reforms to make.

trade presses

Mainstream commercial presses (as opposed to academic presses) that publish books meant for general audiences.

tradition

Beliefs and customs that people justify based on the fact that they have been followed for a long time.

transcript

A complete written copy of a recorded interview or focus group session that contains each word that is spoken on the recording and notes who spoke which words.

treatment

The manipulation of an independent variable in an experiment, usually a specific intervention or factor whose impact on the dependent variable is what the researcher wishes to measure. (Also called the *stimulus*.)

triangulate

To use one research method to evaluate or extend the findings derived from another method.

triangulation

Using one research method to evaluate or extend the findings derived from another method.

trustworthiness

A criterion for assessing the rigor of qualitative research that speaks to the overall truthfulness and usefulness of the results of a study. There are three dimensions of trustworthiness: *credibility*, *dependability*, and *confirmability*.

two-by-two tables

Tables with two rows and two columns that describe how variation across two dimensions leads to distinct outcomes. Researchers use two-by-two tables (and similar diagrams) to illuminate how the categories in a typology relate to one another.

typologies

Classification systems with distinct categories (*types*) that capture different varieties of a given phenomenon.

unit of analysis

The class of phenomena (e.g., individuals, groups, objects, societies) that researchers want to learn about through their research.

units of observation

The class of phenomena (e.g., individuals, groups, objects, societies) that researchers can actually observe for their study, which may also be their unit of analysis or may be another unit that provides indirect knowledge of their unit of analysis.

unstructured interviews

Interviews in which no list of questions exists. In unstructured interviews, the researcher improvises questions as they see fit regarding a particular topic they want to learn about.

user experience research

A type of research that seeks to better understand the perspective of the *end user*, the person who will ultimately use the company's product or service. (Also known as UX or *design research*.)

validity

The accuracy of a measure (also called *construct validity*). A measure is said to be valid if it accurately reflects the meaning of the concept under study.

value

A specific measurement or observed level of a variable.

variable

A quantity or characteristic that can vary. Although scientists often use this term interchangeably with concept, a variable is technically the operational definition of a *concept*—the way the abstract concept is measured in the real world.

venue-based recruitment

Recruiting potential interviewees by posting about your study in offline or online spaces.

weighted sample

A sample that has been adjusted to better reflect the distribution of relevant characteristics in the population.

weighting

In sampling, adjusting for how much particular cases contribute to the statistics for a sample.

WEIRD societies

Societies that fall into the categories of "Western, educated, industrialized, rich, and democratic"—which, given inequalities in where scientific research occurs, tend to be where samples for many studies are drawn.

white paper

A report that presents an organization's take on an issue, often meant to spark discussion and debate among policymakers, the public, or other audiences.

working definition

A draft definition for a particular concept that the researcher uses at the initial stages of a study to help guide their research.

working research question

An initial research question that a researcher develops to guide the design of their study and its early phases.

Credits

Note: Attributions for photographs are listed in the order in which they appear in the chapter. If not otherwise stated, material is used under the doctrine of fair use.

1. Introduction

This chapter includes original material, as well as material taken from *Social Science Research: Principles, Methods, and Practices* (2019), by Anol Bhattacherjee, and *Principles of Sociological Inquiry: Qualitative and Quantitative Methods* (2012), by Amy Blackstone. Both sources are licensed under a CC BY-NC-SA 4.0 license.

Images from beginning of chapter:

Grayscale photo of people at a market, by Danilo Ugaddan, is licensed under a Pexels license. The adapted image, by Bizhan Khodabandeh, is licensed under a CC BY-NC-SA 4.0 license. Modifications included cropping and recoloring.

People watching dancers in traditional clothing, by Mio Advincula, is licensed under a Pexels license.

Engraved portrait of Isaac Newton, by an unknown artist, is public domain and available at Wikimedia Commons. Modifications included recoloring the image.

1.1. The World according to Science

Still image of *Leave it to Beaver*, © Universal, is available at IMDb.

Still image of *The Simpsons*, © Walt Disney Studios, is available at CNET.

Still image of *A Christmas Story* (1983), © Time Warner, is available at *PhillyVoice*.

Roast ham, by Tim Sackton, is licensed under a CC BY-SA 2.0 license and is available at Flickr.

Optical-glass triangular prism, by Dobromir Hristov, is licensed under a Pexels license.

1.2. Sociology as a Social Science

Man in red robe sitting beside green metal fence, by Minh N, is licensed under a Pexels license.

Post by @neiltyson, © Neil deGrasse Tyson, is available at X.

Book cover on the left is from *Cut Loose: Jobless and Hopeless in an Unfair Economy* (2015), by Victor Tan Chen, © University of California Press.

Book cover in the middle is from *A Company of One: Insecurity, Independence, and the New World of White-Collar Unemployment* (2011), by Carrie M. Lane, © Cornell University Press.

Book cover on the right is from *The End of the Line: Lost Jobs, New Lives in Postindustrial America* (1997), by Kathryn Marie Dudley, © University of Chicago Press.

Two girls using laptops surrounded by their classmates, by Max Fischer, is licensed under a Pexels license.

1.3. Design and Goals of This Textbook

People working on building during daytime, by Josue Isai Ramos Figueroa, is licensed under an Unsplash license.

2. Using Sociology in Everyday Life

This chapter includes original material, as well as material taken from *Social Science Research: Principles, Methods, and Practices* (2019), by Anol Bhattacherjee, and *Principles of Sociological Inquiry: Qualitative and Quantitative Methods* (2012), by Amy Blackstone. Both sources are licensed under a CC BY-NC-SA 4.0 license.

Images from beginning of chapter:

People walking near a white building, by Sachith Ravishka Kodikara, is licensed under a Pexels license. The adapted image, by Bizhan Khodabandeh, is licensed under a CC BY-NC-SA 4.0 license. Modifications included cropping and recoloring.

Graphs of performance analytics on a laptop screen, by Luke Chesser, is licensed under an Unsplash license.

2.1. Sociologists as Scholars

Man standing in front of people sitting on red chairs, by the ICSA, is licensed under a Pexels license.

M. Paola Ometto and Carly Offidani-Bertrand presenting at the 2022 Society for the Advancement of Socio-Economics (SASE) conference in Amsterdam, by Victor Tan Chen, is licensed under a CC BY-NC-SA 4.0 license.

"Thesis Defense," by Randall Munroe, is licensed under a CC BY-SA 2.0 license and is available at XKCD.

2.2. Sociologists as Market Researchers

Women colleagues gathered inside a conference room, by Christina Morillo, is licensed under a Pexels license.

Woman looking at a computer monitor, by the Politécnico Grancolombiano Departamento de Comunicaciones, is licensed under a CC BY-SA 2.0 license and is available at Flickr.

2.3. Sociologists as Policy Advisors and Evaluators

Photo of Patrick Moynihan, by John Mathew Smith, is licensed under a CC BY-SA 2.0 license and is available at Wikimedia Commons.

Ronald Reagan's presidential portrait (1981), is public domain and is available at Wikimedia Commons.

Portrait of Michelle Obama in the Green Room of the White House, by Chuck Kennedy, is public domain and is available at Wikimedia Commons.

Report cover, © Data Foundation, is from one of its publications. *Source*: Williams, Stacey, Lori Gonzalez, Phylicia Bediako, Minnie McMillian, Alice Quach, and Austin Hepburn. 2022. "Understanding Policing in America." Washington, DC: Data Foundation.

Headquarters building of the U.S. Census Bureau in Suitland, Maryland, is public domain and is available at Wikimedia Commons.

Logos in the collage are © American Sociological Association, Idealist.org, PublicServiceCareers.org, USAJobs, and the Young Nonprofit Professionals Network.

CWA Local 9119 (AFL-CIO) on strike, by Amgine, is licensed under a CC BY-SA 1.0 license and is available at Wikimedia Commons.

Logos in the collage are © Brookings Institution, Cato Institute, Center for American Progress, Center on Budget and Policy Priorities, Economic Policy Institute, Heritage Foundation, and Manhattan Institute.

Report cover, © Pew Charitable Trusts, is from a Pew Research Center publication. *Source*: Nadeem, Reem. 2022. "Modest Declines in Positive Views of 'Socialism' and 'Capitalism' in U.S. Pew Research Center." Washington, DC: Pew Research Center.

2.4. Sociologists as Clinicians and Public Health Researchers

Photo of Jane Addams has no copyright restrictions and is available from the George Grantham Bain Collection at the U.S. Library of Congress, Prints and Photographs Division [LC-B2- 6531-14].

Works Progress Administration poster advertising Hull House community workshop (1938) is public domain and is available at the U.S. Library of Congress, Prints and Photographs Division [LC-USZC2-5215].

2.5. Sociologists as Activists and Change Agents

Photo of Martin Luther King Jr. has no copyright restrictions and is available at Wikimedia Commons.

Photo of W. E. B. Du Bois (circa 1907) is public domain and is available at Wikimedia Commons.

Map of the distribution of African American inhabitants of the Seventh Ward of Philadelphia, from *The Philadelphia Negro* (1899) by W. E. B. DuBois, is public domain and is available at Wikimedia Commons.

Photo of Sarah Dewees is used with permission.

2.6. Sociologists as Professionals and Citizens

A woman reporting beside the van, by cottonbro studio, is licensed under a Pexels license.

Camels advertisement (1946) is available at Stanford Research into the Impact of Tobacco Advertising (SRITA).

3. The Role of Theory in Research

This chapter includes original material, as well as material taken from *Social Science Research: Principles, Methods, and Practices* (2019), by Anol Bhattacherjee, and *Principles of Sociological Inquiry: Qualitative and Quantitative Methods* (2012), by Amy Blackstone. Both sources are licensed under a CC BY-NC-SA 4.0 license.

Images from beginning of chapter:

Close-up photo of survey spreadsheet, by Lukas, is licensed under a Pexels license. The adapted image, by Bizhan Khodabandeh, is licensed under a CC BY-NC-SA 4.0 license. Modifications included cropping and recoloring.

Colorful dolls on display, by Marc Tollas, is licensed under a Pexels license.

3.1. What Do We Mean by "Theory"?

Shallow-focus photo of woman smiling while holding pen and paper, by Andrea Piacquadio, is licensed under a Pexels license.

3.2. Levels of Analysis

Macro, meso, and micro levels of analysis, by Bizhan Khodabandeh, is licensed under a CC BY-NC-SA 4.0 license.

Homeless man sleeping near a cardboard sign, by Timur Weber, is licensed under a Pexels license.

Photo of Chinyere Osuji is used with permission.

3.3. The Components of Theories

Person riding a horse, by Brett Sayles, is licensed under a Pexels license.

Family having a picnic, by Kampus Production, is licensed under a Pexels license.

Woman and girl talking while lying on bed, by Ketut Subiyanto, is licensed under a Pexels license.

People clapping their hands while seated on a sofa, by cottonbro studio, is licensed under a Pexels license.

The theoretical and empirical levels of research, by Bizhan Khodabandeh, is licensed under a CC BY-NC-SA 4.0 license.

The relationships between concepts, by Bizhan Khodabandeh, is licensed under a CC BY-NC-SA 4.0 license.

Close-up photo of old machinery, by Tima Miroshnichenko, is licensed under a Pexels license.

Charts illustrating spurious correlations, by Tyler Vigen, is licensed under a CC BY 4.0 license and is available at Spurious Correlations.

Screenshot of S.T.A.L.K.E.R.: *Call of Pripyat* is licensed under a GNU Free Documentation license and is available from GSC Game World via Wikimedia Commons.

American pitbull terrier puppy lying on the floor, by Joice Rivas, is licensed under a Pexels license.

Screenshot of a *Daily Mail* photo, © Jason Galvin, and an article, "Army Veteran Rescues Bald Eagle Dangling Upside Down from a Rope in 75-Foot Tree by 'Mowing Down the Branches' with his Rifle," by Valerie Edwards, © Daily Mail and General Trust, are available at Reddit.

Hygge book on a blanket, by Stella Rose, is licensed under an Unsplash license.

3.4. Evaluating Theories

Protest against pandemic measures in Liestal, Switzerland, by Kajetan Sumila, is licensed under an Unsplash license.

Portrait of William of Ockham is licensed under a CC BY 4.0 license and is available at Wikimedia Commons.

3.5. Major Sociological Theories and Paradigms

Group huddling on a grass field, by Anastasia Shuraeva, is licensed under a Pexels license.

Man pointing his finger, by Nathan Shively, is licensed under an Unsplash license.

Portrait of Thomas Kuhn, by Davi.trip, is licensed under a CC BY-SA 4.0 license and is available at Wikimedia Commons.

Man with red-dyed hair, a pink shirt, and tinted glasses, by Reymon Reymon, is licensed under a Pexels license.

Photo of Auguste Comte, by Johan Hendrik Hoffmeister, is public domain and is available at Wikimedia Commons.

Photo of Émile Durkheim is public domain and is available at Wikimedia Commons.

Photo of Paul Lazarsfeld, © Bardwell Press, is available at Wikimedia Commons.

Person's left hand against a white surface, by Elena Rabkina, is licensed under an Unsplash license.

Photo of Max Horkheimer, by Barbara Niggl Radloff, is licensed under a CC BY 4.0 license and is available at Wikimedia Commons. Modifications included cropping.

Photo of Nancy Fraser, by Bunnyfrosch, is licensed under CC ShareAlike 3.0 Unported, 2.5 Generic, 2.0 Generic and 1.0 Generic licenses and is available at Wikimedia Commons. Modifications included cropping and recoloring.

3.6. Linking Theory and Data

Photo of Albert Einstein, by Orren Jack Turner, has no known copyright restrictions and is available at Wikimedia Commons. Modifications included cropping.

Photo of Stephen Hawking, by Doug Wheller, is licensed under a CC BY 2.0 license and is available at Wikimedia Commons. Modifications included cropping and recoloring.

4. Research Questions

This chapter includes original material, as well as material taken from *Scientific Inquiry in Social Work* (2018), by Matthew DeCarlo, which is licensed under a CC BY-NC-SA 4.0 license.

Images from beginning of chapter:

Brown wooden bookshelves with books, by Olena Bohovyk, is licensed under a Pexels license. The adapted image, by Bizhan Khodabandeh, is licensed under a CC BY-NC-SA 4.0 license. Modifications included cropping and recoloring.

4.1. Choosing a General Research Topic

People gathered near building holding LGBTQ+ flags, by Gotta Be Worth It, is licensed under a Pexels license.

Model of multiple railroad tracks, by Google DeepMind, is licensed under a Pexels license.

Photo of Joyce Rothschild, by Gabriela León-Pérez, is licensed under a CC BY-NC-SA 4.0 license.

Man holding black DSLR camera, by Suliman Sllehi, is licensed under a Pexels license.

Example of a concept map, by Bizhan Khodabandeh, is licensed under a CC BY-NC-SA 4.0 license.

4.2. Types of Research Questions

Still image of *The SpongeBob SquarePants Movie* (2004), © Nickelodeon, is available at IMDb.

Asian friends with tattoos, by Khoa Võ, is licensed under a Pexels license.

4.3. Posing Explanatory Research Questions

The wheel of science, by Bizhan Khodabandeh, is licensed under a CC BY-NC-SA 4.0 license.

Deduction: from abstract to concrete, by Bizhan Khodabandeh, is licensed under a CC BY-NC-SA 4.0 license.

Rally to stop Asian hate, McPherson Square, Washington, DC (March 21, 2021), by Victoria Pickering, is licensed under a CC BY-NC-ND 2.0 license and is available at Flickr.

Induction: from concrete to abstract, by Bizhan Khodabandeh, is licensed under a CC BY-NC-SA 4.0 license.

The Usual Irish Way of Doing Things, by Thomas Nast, published in *Harper's Weekly* on September 2, 1871, is public domain and is available at Wikimedia Commons.

Screenshot of a Reddit post, "I Had a Science Fair Activity Tonight," © whiteinshiningarmor, is available at Reddit.

4.4. Refining Your Research Question

Multiethnic family with a baby at home, by William Fortunato, is licensed under a Pexels license.

5. Research Design

This chapter includes original material, as well as material taken from *Scientific Inquiry in Social Work* (2018), by Matthew DeCarlo, which is licensed under a CC BY-NC-SA 4.0 license.

Images from beginning of chapter:

Magnifying glass on top of a document, by Anna Nekrashevich, is licensed under a Pexels license. The adapted image, by Bizhan Khodabandeh, is licensed under a CC BY-NC-SA 4.0 license. Modifications included cropping and recoloring.

5.1. What Should My Literature Review Cover?

"Mind the Gap" warning at a train station, by raghavvidya, is licensed under a CC BY-NC-ND 2.0 license and is available at Flickr.

Stack of hardbound books, by cottonbro studio, is licensed under a Pexels license.

5.2. What Sources Should I Use?

Man researching alone at the library, by Tima Miroshnichenko, is licensed under a Pexels license.

Meme about Reviewer 2, © xixoxixa, is available at Reddit.

Wikipedia logo, by Nohat (concept by Paullusmagnus), is licensed under a CC BY-SA 3.0 license.

5.3. How Do I Find Relevant Sources?

Low-angle photo of white bookshelves, by CHUTTERSNAP, is licensed under an Unsplash license.

5.4. How Do I Read and Evaluate Sources?

Post by @SciBry, © Brian Gaensler, is available at X.

Man working on laptop and taking notes, by Ono Kosuki, is licensed under a Pexels license.

Literature review matrix, @ Kelsey Cheshire, is available at the VCU Libraries website.

Signpost, by Zoe Davidson, is licensed under a CC BY-NC-ND 2.0 license and is available at Flickr.

5.5. How Do I Write a Literature Review?

Assorted beans placed in rows on white fabric, by Mariam Antadze, is licensed under a Pexels license.

6. Sampling

This chapter includes original material, as well as material taken from *Social Science Research: Principles, Methods, and Practices* (2019), by Anol Bhattacherjee, and *Principles of Sociological Inquiry: Qualitative and Quantitative Methods* (2012), by Amy Blackstone. Both sources are licensed under a CC BY-NC-SA 4.0 license.

Images from beginning of chapter:

Peeled tangerine, by Karolina Kaboompics, is licensed under a Pexels license. The adapted image, by Bizhan Khodabandeh, is licensed under a CC BY-NC-SA 4.0 license. Modifications included cropping and recoloring.

Photomontage of faces, by Gerd Altmann, is licensed under a Pixabay license.

6.1. Units of Analysis

Man with a backpack looking through binoculars, by Rory McKeever, is licensed under an Unsplash license.

Units of analysis, populations and samples, by Bizhan Khodabandeh, is licensed under a CC BY-NC-SA 4.0 license.

Person using a magnifying glass on a map, by lil artsy, is licensed under a Pexels license.

6.2. Populations versus Samples

People sitting on stadium seats, by Anna Sullivan, is licensed under an Unsplash license.

Man standing in front of group of men, by Austin Distel, is licensed under an Unsplash license.

Person flipping through the pages of a book, by Brittany Colette, is licensed under an Unsplash license.

Target populations and sampling frames, by Bizhan Khodabandeh, is licensed under a CC BY-NC-SA 4.0 license.

Respondents versus nonrespondents, by Bizhan Khodabandeh, is licensed under a CC BY-NC-SA 4.0 license.

6.3. Probability Sampling

Kid holding candies, by Patrick Fore, is licensed under an Unsplash license.

Left hand playing with dice, by Guillermo Velarde, is licensed under an Unsplash license.

Systematic sampling, by Bizhan Khodabandeh, is licensed under a CC BY-NC-SA 4.0 license.

An example of periodicity when using systematic sampling, by Bizhan Khodabandeh, is licensed under a CC BY-NC-SA 4.0 license.

Examples of proportionate and disproportionate stratified sampling: racial and ethnic groups in Richmond, Virginia, by Bizhan Khodabandeh, is licensed under a CC BY-NC-SA 4.0 license.

An example of cluster sampling: surveying public librarians in Virginia, by Bizhan Khodabandeh, is licensed under a CC BY-NC-SA 4.0 license.

6.4. Nonprobability Sampling

Six assorted-flavor donuts, by Heather Ford, is licensed under an Unsplash license.

People queuing beside a Louis Vuitton store, by Melanie Pongratz, is licensed under an Unsplash license.

Snowball sampling, by Bizhan Khodabandeh, is licensed under a CC BY-NC-SA 4.0 license.

Example of quota sampling, by Bizhan Khodabandeh, is licensed under a CC BY-NC-SA 4.0 license.

Harry S. Truman holding the *Chicago Daily Tribune*, by Byron H. Rollins, is public domain and is available at Wikimedia Commons.

The Man burns at Burning Man 2014, by torroid, is licensed under a CC BY-SA 2.0 license and is available at Flickr.

Flight attendants standing in the aisle of an airplane, by Vinh Lâm, is licensed under a Pexels license.

6.5. Questions to Ask about Samples

"Sampling Bias," by Jono Hey, is licensed under a CC BY-NC 4.0 license and is available at *Sketchplanations*.

Post by @elonmusk, © Elon Musk, is available at X.

7. Measuring the Social world

This chapter includes original material, as well as material taken from *Scientific Inquiry in Social Work* (2018), by Matthew DeCarlo, which is licensed under a CC BY-NC-SA 4.0 license.

Images from beginning of chapter:

Earth from afar, by NASA, is public domain and is available at the NASA Earth Observatory. The adapted image, by Bizhan Khodabandeh, is licensed under a CC BY-NC-SA 4.0 license. Modifications included cropping and recoloring.

The conceptualization-operationalization-measurement (COM) process, by Bizhan Khodabandeh, is licensed under a CC BY-NC-SA 4.0 license.

7.1. Operationalizing a Concept

Still image of the movie *The Marine* (2006), © 20th Century Fox, is available at Wikimedia Commons.

Cleaner holding a bag with plastic bottles, by Xavier Messina, is licensed under a Pexels license.

Photo of Mollie Orshansky is public domain and is available from the Social Security Administration History Archives via Wikimedia Commons.

Book cover is from *The Bell Curve: Intelligence and Class Structure in American Life* (1996), by Richard J. Herrnstein and Charles Murray, © Simon & Schuster and Princeton University Press.

7.2. Deciding on the Correct Level of Measurement

German shepherd lying on the ground, by Summer Stock, is licensed under a Pexels license. Yorkshire terrier in a party hat, by Sam Lion, is licensed under a Pexels license. The adapted image, by Bizhan Khodabandeh, is licensed under a CC BY-NC-SA 4.0 license. Modifications included cropping.

Ordinal level of measurement example: perceived health, by Bizhan Khodabandeh, is licensed under a CC BY-NC-SA 4.0 license.

The hierarchy of levels of measurement, by Bizhan Khodabandeh, is licensed under a CC BY-NC-SA 4.0 license.

Temperature scales, by Emeka Udenze, is licensed under a CC0 1.0 Universal license and is available at Wikimedia Commons.

7.3. Indexes and Scales: Measures with Multiple Items

Image of the Social Progress Index component-level framework, © Social Progress Imperative, is from a report. *Source*: Stern, Scott, Jaromir Harmacek, Petra Krylova, and Mohamed Htitich. 2022. "2022 Social Progress Index Methodology Report." Washington, DC: Social Progress Imperative.

7.4. Measurement Quality

Assorted wine bottles, by Chris F, is licensed under a Pexels license.

Couple relaxing in a jacuzzi, by Ron Lach, is licensed under a Pexels license.

Comparing reliability and validity, by Bizhan Khodabandeh, is licensed under a CC BY-NC-SA 4.0 license.

Man washing his hands, by Ketut Subiyanto, is licensed under a Pexels license.

Photo of L. Douglas Wilder is public domain and is available from the National Institutes of Health via Wikimedia Commons. Modifications included cropping and recoloring.

Photo of Tom Bradley, by Bob Chamberlin, is licensed under a CC BY 4.0 license and is available from the *Los Angeles Times* via the UCLA Library. Modifications included cropping and recoloring.

Women holding Trump banners, by Michael Anthony, is licensed under a Pexels license.

Photo of Matthew Desmond is public domain and available from the U.S. Library of Congress via Wikimedia Commons.

8. Ethics

This chapter includes original material, as well as material taken from *Scientific Inquiry in Social Work* (2018), by Matthew DeCarlo, and *Principles of Sociological Inquiry: Qualitative and Quantitative Methods* (2012), by Amy Blackstone. Both sources are licensed under a CC BY-NC-SA 4.0 license.

Images from beginning of chapter:

Balance scale on a table, by Katrin Bolovtsova, is licensed under a Pexels license. The adapted image, by Bizhan Khodabandeh, is licensed under a CC BY-NC-SA 4.0 license. Modifications included cropping and recoloring.

Guard with sunglasses, by the Stanford Prison Experiment, is used with permission for noncommercial purposes and is available at PrisonExp.org.

Rebellious prisoners, by the Stanford Prison Experiment, is used with permission for noncommercial purposes and is available at PrisonExp.org.

Prisoner 819 leaves the study, by the Stanford Prison Experiment, is used with permission for noncommercial purposes and is available at PrisonExp.org.

8.1. Subject to Harm: Coercion, Deception, and Other Risks for Participants

African American male being tested and treated during the Tuskegee Study of Untreated Syphilis in the Negro Male is public domain and is available from the U.S. Centers for Disease Control and Prevention via Wikimedia Commons.

Romani victim of Nazi medical experiments is public domain and is available from the U.S. National Archives and Records Administration via the U.S. Holocaust Memorial Museum.

Still image of the movie *Wormwood* (2017), © Netflix, is available at the *New York Review*.

Page from the declassified documents of Project MKUltra is public domain and available at Wikimedia Commons.

Defendants in the dock at the Nuremberg war crimes tribunal, by Raymond D'Addario, is public domain and available at Wikimedia Commons.

Sketch of Ted Kaczynski, by Jeanne Boylan, is public domain and available from the U.S. Federal Bureau of Investigation via NPR.

Still image from *The Simpsons* episode "Family Therapy" (1989), © Walt Disney Company, is available at IMDb.

Book cover is from *Tearoom Trade: Impersonal Sex in Public Spaces* (2008), by Laud Humphreys, © Routledge.

8.2. The Belmont Principles: Respect for Persons,

Beneficence, and Justice

Cover page of *The Belmont Report* (1979), by the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, is public domain and available at the University of North Texas Digital Library.

Logo of Cambridge Analytica is licensed under a CC BY-SA 4.0 license and is available at Wikimedia Commons.

1956 mugshot of James "Whitey" Bulger is public domain and is available from USP Atlanta via Wikimedia Commons.

2011 mugshot of Bulger is public domain and is available from the U.S. Marshals Service via Wikimedia Commons.

Photo of Ronald Kessler and Mark Felt is licensed under a CC BY-SA 4.0 license and is available at Wikimedia Commons.

Animal Liberation Front Israel protest in Tel Aviv (September 8, 2016), by Roee Shpernik, is licensed under a CC BY-SA 4.0 license and is available at Wikimedia Commons.

Meme of Donald Trump, adapted from an interview by Jonathan Swan, © *Axios*, is available at *Know Your Meme*.

8.3. Regulating Research: Institutional Review Boards and Professional Norms

Screenshot of an institutional review board website, © Virginia Commonwealth University (VCU), is available at the VCU website.

Logo, © American Sociological Association (ASA), is available at the ASA website.

8.4. Veracity and Objectivity in Research

Still image from the movie *Inside Job* (2010), © Sony Pictures Classics, is available at IMDb.

The Terrible Inspection (circa 1947), by Huang Rong-can, is public domain and is available at Wikimedia Commons.

Prisoners facing the wall, by the Stanford Prison Experiment, is used with permission for noncommercial purposes and is available at PrisonExp.org.

Photo of Didem Danış is used with permission.

9. Ethnography

This chapter includes original material, as well as material taken from *Social Science Research: Principles, Methods, and Practices* (2019), by Anol Bhattacherjee, and *Principles of Sociological Inquiry: Qualitative and Quantitative Methods* (2012), by Amy Blackstone. Both sources are licensed under a CC BY-NC-SA 4.0 license.

Images from beginning of chapter:

Close-up photo of a person, by Marina Vitale, is licensed under an Unsplash license. The adapted image, by Bizhan Khodabandeh, is licensed under a CC BY-NC-SA 4.0 license. Modifications included cropping and recoloring.

A couple fighting, by Timur Weber, is licensed under a Pexels license.

Police officer standing near bare trees, by Kindel Media, is licensed under a Pexels license.

9.1. When and How to Use Ethnography

Two senior Inuits talking outdoors, by Александр Велигура, is licensed under a Pexels license.

Cockfighting (circa 1958) is public domain and is available from the Indonesian Ministry of Information via Wikimedia Commons.

Person holding meat and a knife, by Leonardo Luz, is licensed under a Pexels license.

The participation continuum, by Bizhan Khodabandeh, is licensed under a CC BY-NC-SA 4.0 license.

Still image from the *Looney Tunes* episode "Mad as a Mars Hare," © Warner Bros., is available at Wikimedia Commons.

Still image from the *South Park* episode "Going Native" (2012), © Comedy Central, is available at IMDB.

9.2. Going into the Field

AddUpDog letterhead, by Elijah Kai Chen, is licensed under a CC BY-NC-SA 4.0 license.

Sorority Bid Day, by Henderson State University, is licensed under a CC BY-NC 2.0 license and is available at Flickr.

The disclosure continuum, by Bizhan Khodabandeh, is licensed under a CC BY-NC-SA 4.0 license.

Survivalist with a bow and backpack standing in a mountain pasture, by Jordan Bergendahl, is licensed under a Pexels license.

Photo of Margaret Mead has no known copyright restrictions and is available from the Smithsonian Institution via Flickr.

Photo of Amy L. Best is used with permission.

9.3. Field Jottings and Field Notes

Book pages are from *Cut Loose: Jobless and Hopeless in an Unfair Economy* (2015), by Victor Tan Chen, © University of California Press.

Women using mobile phones, by Brett Sayles, is licensed under a Pexels license.

Woman in a checkered suit conducting an interview, by Tima Miroshnichenko, is licensed under a Pexels license.

Man beside a truck, by the World Sikh Organization of Canada, is licensed under a Pexels license.

10. In-Depth Interviewing

This chapter includes original material, as well as material taken from *Scientific Inquiry in Social Work* (2018), by Matthew DeCarlo, which is licensed under a CC BY-NC-SA 4.0 license.

Images from beginning of chapter:

Woman talking to a radio host, by George Milton, is licensed under a Pexels license. The adapted image, by Bizhan Khodabandeh, is licensed under a CC BY-NC-SA 4.0 license. Modifications included cropping and recoloring.

Sara Rathod interviews Linda Wertheimer after a debate Wertheimer moderated at Wellesley College, by Graham Milldrum, is licensed under a CC BY 2.0 license and is available at Flickr.

10.1. When Should In-Depth Interviewing Be Used?

Crying toddler holding a banana, by Quinn Dombrowski, is licensed under a CC BY-SA 2.0 license and is available at Flickr.

10.2. Creating Interview Guides

"Prefer Open-Ended Questions," by Jono Hey, is licensed under a CC BY-NC 4.0 license and is available at *Sketchplanations*.

10.3. Preparing for In-Depth Interviews

Interviewing workers at the Hongcheon pinenut processing factory, by Ramon Carrillo, is licensed under a CC BY-NC-SA 2.0 license and is available from the Asia-Pacific Forest Communicator Network via Flickr.

10.4. Conducting In-Depth Interviews

"Don't Fill the Silence," by Jono Hey, is licensed under a CC BY-NC 4.0 license and is available at *Sketchplanations*.

Peer editing during a writer's workshop, by Tim Lauer, is licensed under a CC BY-NC 2.0 license and is available at Flickr.

Photo of Allison Pugh, by Dan Addison, is used with permission.

10.5. Focus Groups

Grayscale photo of man in a black snapback hat, by Marcus Bellamy, is licensed used under an Unsplash license.

11. Qualitative Data Analysis

This chapter includes original material, as well as material taken from *Scientific Inquiry in Social Work* (2018), by Matthew DeCarlo, and *Social Science Research: Principles, Methods, and Practices* (2019), by Anol Bhattacherjee,. Both sources are licensed under a CC BY-NC-SA 4.0 license.

Images from beginning of chapter:

Scattered letter blocks, by Suzy Hazelwood, is licensed under a Pexels license. The adapted image, by Bizhan Khodabandeh, is licensed under a CC BY-NC-SA 4.0 license. Modifications included cropping and recoloring.

11.1. Coding Qualitative Data

Assorted colored stones in wooden containers, by julie aagaard, is licensed under a Pexels license.

Logos in the collage are © ATLAS.ti Scientific Software Development GmbH (ATLAS.ti), SocioCultural Research Consultants (Dedoose), VERBI Software (MaxQDA), Lumivero (NVivo), and Provalis Research (QDA Miner).

Grayscale photo of a man smashing bricks, by Alexa Popovich, is licensed under a Pexels license. Modifications included cropping.

Brick layer in Tel Aviv, by Zoltan Kluger, is public domain and is available at Wikimedia Commons. Modifications included cropping.

Article page, © Oxford Academic, is from a journal article. *Source*: Chen, Victor Tan, and Jesse Goldstein. 2024. "The Theatre of Entrepreneurship: Learning to Perform the Speculative Self in University Entrepreneurship Programs." *Social Problems* 71(1):203–19. doi: 10.1093/socpro/spac008.

Photo of Leslie Martin is courtesy of the University of Mary Washington.

11.2. Creating Arguments out of Codes

Representation of the evolution of the universe over 13.77 billion years, by NASA/WMAP Science Team, is public domain and is available at Wikimedia Commons.

Meme with a still image from the movie *Mean Girls* (2004), © Paramount, is available at Fandom.

Two-by-two table, © Wiley, is from a journal article. *Source*: Doody, Sean, Victor Tan Chen, and Jesse Goldstein. 2016. "Varieties of Entrepreneurial Capitalism: The Culture of Entrepreneurship and Structural Inequalities of Work and Business Creation." *Sociology Compass* 10(10):858–76. doi: 10.1111/soc4.12407.

Diagram representing a typology, by Victor Tan Chen, is based on material from a handbook chapter and is licensed under a CC BY-NC-SA 4.0 license. *Source*: Chen, Victor Tan. 2021. "The Mirage of Meritocracy and the Morality of Grace." Pp. 58–72 in *The Routledge Handbook on the American Dream*. Vol. 1, edited by R. C. Hauhart and M. Sardoč. New York: Routledge.

Concept map representing relationships among variables, by Victor Tan Chen, is based on material from a journal article and is licensed under a CC BY-NC-SA 4.0 license. *Source*: Dinesen, Peter Thisted, Merlin Schaeffer, and Kim Mannemar Sønderskov. 2020. "Ethnic Diversity and Social Trust: A Narrative and Meta-Analytical Review." *Annual Review of Political Science* 23(1):441–65. doi: 10.1146/annurev-polisci-052918-020708.

11.3. Writing Up Qualitative Results

Close-up photo of a burger, by Vilnis Husko, is licensed under a Pexels license. Modifications included cropping.

Twin girls smiling, by cottonbro studio, is licensed under a Pexels license.

Blockquotes, © Sage, are from a journal article. *Source*: Ravenelle, Alexandrea J., Abigail Newell, and Ken Cai Kowalski. 2021. "'The Looming, Crazy Stalker Coronavirus': Fear Mongering, Fake News, and the Diffusion of Distrust." *Socius* 7:23780231211024776. doi: 10.1177/23780231211024776.

Family preparing food in the kitchen, by August de Richelieu, is licensed under a Pexels license.

Book excerpts are from *Cut Loose: Jobless and Hopeless in an Unfair Economy* (2015), by Victor Tan Chen, © University of California Press.

12. Experiments

This chapter includes original material, as well as material taken from *Scientific Inquiry in Social Work* (2018), by Matthew DeCarlo, *Social Science Research: Principles, Methods, and Practices* (2019), by Anol Bhattacherjee, and *Research Methods in Psychology* (2015), by Rajiv S. Jhangiani and I-Chant A. Chiang. All sources are licensed under a CC BY-NC-SA 4.0 license.

Images from beginning of chapter:

Two test tubes, by Martin Lopez, is licensed under a Pexels license. The adapted image, by Bizhan Khodabandeh, is licensed under a CC BY-NC-SA 4.0 license. Modifications included cropping and recoloring.

Still image from the *Mythbusters* episode "President's Challenge," © Discovery Channel, is available at NPR.

12.1. Pre-experimental Designs: What You Should Not Do

Boy in a white lab coat holding an eye dropper, by Clint Patterson, is licensed under an Unsplash license.

AddUpDog logo, by Elijah Kai Chen, is licensed under a CC BY-NC-SA 4.0 license.

One-group posttest-only design, by Bizhan Khodabandeh, is licensed under a CC BY-NC-SA 4.0 license.

One-group pre-post design, by Bizhan Khodabandeh, is licensed under a CC BY-NC-SA 4.0 license.

Two-group pretest–posttest design, by Bizhan Khodabandeh, is licensed under a CC BY-NC-SA 4.0 license.

Apples and oranges on a plaid blanket, by Anastasiya Romanova, is licensed under an Unsplash license.

Ad for participants, by the Stanford Prison Experiment, is used with permission for noncommercial purposes and is available at PrisonExp.org.

12.2. Randomized Controlled Designs: What You Should Do

A classic experiment: two-group randomized pre-post design, by Bizhan Khodabandeh, is licensed under a CC BY-NC-SA 4.0 license.

Still image from the *Star Trek* episode "The City on the Edge of Forever" (1967), © CBS/Paramount, is available at *Trek Report*.

Two-group randomized pre-post design with scores (math app example), by Bizhan Khodabandeh, is licensed under a CC BY-NC-SA 4.0 license.

12.3. Persistent Validity Problems: What You Still Need to Avoid

Pygmalion Adoring His Statue (1717), by Jean Raoux, is public domain and is available at Wikimedia Commons.

Pregnant woman typing on a laptop, by Pavel Danilyuk, is licensed under a Pexels license.

White pill on a yellow surface, by Karolina Grabowska, is licensed under a Pexels license.

Young man working on a laptop and writing notes, by Michael Burrows, is licensed used under a Pexels license.

Two-group randomized posttest-only design, by Bizhan Khodabandeh, is licensed under a CC BY-NC-SA 4.0 license.

12.4. Field Experiments: External Validity at the Cost of Control

Police officer putting handcuffs on another person, by Kindel Media, is licensed under a Pexels license.

Outcomes at 24 years or older for children who moved before they were 13: an analysis of Moving to Opportunity data, by Bizhan Khodabandeh, is based on material from a news article and journal article and is licensed under a CC BY-NC-SA 4.0 license. *Graphic adapted from source*: Wolfers, Justin. 2015. "Why the New Research on Mobility Matters: An Economist's View." *New York Times*, May 4. *Data updated from source:* Chetty, Raj, Nathaniel Hendren, and Lawrence F. Katz. 2016. "The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment." *American Economic Review* 106(4):855–902. doi: 10.1257/aer.20150572.

Handshake, by Thirdman, is licensed under a Pexels license.

Appointments obtained for mental health therapy by the class, race, and gender of the help-seeker: an audit study, by Bizhan Khodabandeh, is based on material in a journal article and is licensed under a CC BY-NC-SA 4.0 license. *Source*: Kugelmass, Heather. 2016. "'Sorry, I'm Not Accepting New Patients': An Audit Study of Access to Mental Health Care." *Journal of Health and Social Behavior* 57(2):1–16. doi: 10.1177/0022146516647098.

The effect of a criminal record on black and white job applicants: an audit study, by Bizhan Khodabandeh, is based on material in a journal article and is licensed under a CC BY-NC-SA 4.0 license. *Source*: Pager, Devah. 2003. "The Mark of a Criminal Record." *American Journal of Sociology* 108(5):937–75. doi: 10.1086/374403.

Sitting woman in a striped jacket shaking hands with someone standing, by the RDNE Stock Project, is licensed under a Pexels license.

12.5. Quasi-experiments: Approximating True Experiments with Observational Data

Twin girls holding red and yellow skateboards, by cottonbro studio, is licensed under a Pexels license.

Nonequivalent control group design, by Bizhan Khodabandeh, is licensed under a CC BY-NC-SA 4.0 license.

Graph illustrating the effects of the CESSP Scholarship Program on school attainment, © Deon Filmer and Norbert Schady, is from a working paper. *Source*: Filmer, Deon, and Norbert Schady. 2009. "School Enrollment, Selection, and Test Scores." Policy Research Working Paper No. 4998, World Bank, Washington, DC.

Demolished residential buildings, by Franklin Peña Gutierrez, is licensed under a Pexels license.

The logic of the differences-in-differences method in a minimum-wage study, by Bizhan Khodabandeh, is based on material from a journal article and is licensed under a CC BY-NC-SA 4.0 license. *Source*: Card, David, and Alan Krueger. 1994. "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania." *American Economic Review* 84(4):772–93.

Barack Obama delivering a speech at the Washington National Cathedral, by Jose A. Torres Jr., is public domain and is available from the U.S. Department of Defense via Wikimedia Commons.

Automobiles destroyed by the Oklahoma City bombing, by Preston Chasteen, is public domain and is available at Wikimedia Commons.

One-group time-series design, by Bizhan Khodabandeh, is licensed under a CC BY-NC-SA 4.0 license.

Time-series outcome when the treatment has an effect, by Bizhan Khodabandeh, is licensed under a CC BY-NC-SA 4.0 license.

Graph (Figure 1a) illustrating the Oklahoma County general fertility rate over time, © Springer, is from a journal article. *Source*: Rodgers, Joseph Lee, Craig A. St John, and Ronnie Coleman. 2005. "Did Fertility Go up after the Oklahoma City Bombing? An Analysis of Births in Metropolitan Counties in Oklahoma, 1990-1999." *Demography* 42(4):675–92. doi: 10.1353/dem.2005.0034.

Multiple time-series design with two comparison groups, by Bizhan Khodabandeh, is licensed under a CC BY-NC-SA 4.0 license.

12.6. Experiments: Our Best Way of Inferring Causality, but Far from Foolproof

Two girls eating cookies, by cottonbro studio, is licensed under a Pexels license.

13. Surveys

This chapter includes original material, as well as material taken from *Social Science Research: Principles, Methods, and Practices* (2019), by Anol Bhattacherjee, *Principles of Sociological Inquiry: Qualitative and Quantitative Methods* (2012), by Amy Blackstone, and *Research Methods in Psychology* (2015), by Rajiv S. Jhangiani and I-Chant A. Chiang. All sources are licensed under a CC BY-NC-SA 4.0 license.

Images from beginning of chapter:

Woman reading papers, by cottonbro studio, is licensed under a Pexels license. The adapted image, by Bizhan Khodabandeh, is licensed under a CC BY-NC-SA 4.0 license. Modifications included cropping and recoloring.

Photo of Shirley Chisholm, by Thomas J. O'Halloran, has no known copyright restrictions and is available at the U.S. Library of Congress, Prints and Photographs Division [LC-U9- 25383-33].

13.1. The Strengths and Weaknesses of Survey Research

Cover of U.S. *News & World Report*'s 2022 edition of "Best Colleges Ranking," © U.S. News & World Report LP, is available at Wikipedia.

Annoyed woman having a video chat on her smartphone at home, by Alex Green, is licensed under a Pexels license.

13.2. Types of Surveys

SAG-AFTRA strike rally, by Eden, Janine and Jim, is licensed under a CC BY 2.0 license and is available at Flickr.

Section of the 1890 U.S. Census form is public domain and is available from the U.S. Census Bureau.

Letter regarding the 2023 Survey of Doctorate Recipients is public domain and is available from the National Science Foundation.

Question for a survey interviewer, © NORC, is from a 2022 General Social Survey (GSS) questionnaire and is available at the GSS website.

13.3. Designing Effective Questions and Questionnaires

Table summarizing results from three separate but similarly worded questions, © Pew Charitable Trusts, is from a Pew Research Center report. *Source:* Tyson, Alec, and Carroll Doherty. 2013. "Polling on the Deficit: Why Question Order Matters." Washington, DC: Pew Research Center.

Question about family income, © NORC, is from a 2022 GSS questionnaire and is available at the GSS website.

Excerpt of survey for the 2016 U.S. presidential candidates, © Global Tetrahedron, is from an *Onion* article. *Source: The Onion*. 2016. "'The Onion' Is Withholding Our Endorsement for President until Both Candidates Respond to Our Questionnaire." *The Onion*, October 12.

Double-barreled response options, by Bizhan Khodabandeh, is licensed under a CC BY-NC-SA 4.0 license.

Excerpt of survey questions in a matrix format, © Mexican Migration Project, is available via Princeton University.

Excerpt of filter questions from the 2016 Survey of Consumer Finances questionnaire is public domain and available from the U.S. Federal Reserve.

14. Quantitative Data Analysis

This chapter includes original material, as well as material taken from *Social Science Research: Principles, Methods, and Practices* (2019), by Anol Bhattacherjee, which is licensed under a CC BY-NC-SA 4.0 license.

Images from beginning of chapter:

Colorful plastic numbers and mathematical symbols, by Karolina Grabowska, is licensed under a Pexels license. The adapted image, by Bizhan Khodabandeh, is licensed under a CC BY-NC-SA 4.0 license. Modifications included cropping and recoloring.

Still image from the *Schoolhouse Rock* episode "Ready or Not Here I Come" (1973), © Scholastic Rock, Inc./ABC, is available at Schoolhouse Rock Fandom.

Logo of SPSS, © IBM, is available at the SPSS website.

14.1. Preparing for Data Analysis

Codebook excerpt from the LIFE Dataset, © Mexican Migration Project, is available via Princeton University.

Levels of measurement, by Bizhan Khodabandeh, is licensed under a CC BY-NC-SA 4.0 license.

Two people's hands touching, by Alexander Grey, is licensed under an Unsplash license.

Close-up photo of a computer mining rig, by panumas nikhomkhai, is licensed under a Pexels license.

Screenshot of a data repository, © Inter-university Consortium for Political and Social Research (ICPSR), is available at the ICPSR website.

Screenshot of the GSS Data Explorer, © NORC, is available at the GSS website.

Excerpt from a 2022 GSS questionnaire, © NORC, is available at the GSS website.

14.2. Univariate Data Analysis: Analyzing One Variable at a Time

Breaking down a frequency table, by Bizhan Khodabandeh, is licensed under a CC BY-NC-SA 4.0 license.

Businessman holding money in his hands, by Andrea Piacquadio, is licensed under a Pexels license.

People on a cliff watching a man dive into the sea, by Kammeran Gonzalez-Keola, is licensed under a Pexels license.

Elderly person holding a stress ball as someone touches their wrist, by Matthias Zomer, is licensed under a Pexels license.

Breakdown of a sample's demographic characteristics (Table 1), © ScienceDirect, is from a journal article. *Source*: Lee, Bob, Yiwei Chen, and Lynne Hewitt. 2011. "Age Differences in Constraints Encountered by Seniors in Their Use of Computers and the Internet." *Computers in Human Behavior* 27(3):1231–37. doi: 10.1016/j.chb.2011.01.003.

Photo of Elyas Bakhtiari is used with permission.

14.3. Bivariate Data Analysis: Crosstabulations and

Chi-Square

Woman in a brown dress praying with her arms raised, by Luis Quintero, is licensed under a Pexels license.

Pumpkins of different sizes, by Nathan J. Hilton, is licensed under a Pexels license.

"Null Hypothesis," by Randall Munroe, is licensed under a CC BY-NC 2.5 license and is available at XKCD.

Breakdown of healthcare behaviors by level of literacy (Table 2), © International Childbirth Education Association, is from a journal article. *Source*: Kimunai, Eunice, Salome Kapella-Mshigeni, and Peter Anderson. 2016. "Mother-to-Child HIV Transmission, Literacy, Ethnicity, Education, and Wealth in Kenya." *International Journal of Childbirth Education* 31(2):8–17.

15. Materials-Based Methods

This chapter includes original material, as well as material taken *Principles of Sociological Inquiry: Qualitative and Quantitative Methods* (2012), by Amy Blackstone, which is licensed under a CC BY-NC-SA 4.0 license.

Images from beginning of chapter:

Tired woman in the library, by Ron Lach, is licensed under a Pexels license. The adapted image, by Bizhan Khodabandeh, is licensed under a CC BY-NC-SA 4.0 license. Modifications included cropping and recoloring.

Athletes at a starting line, by Gaby Lopez, is licensed under a Pexels license.

15.1. Content Analysis

Bond girls compilation, by MachoCarioca, is licensed under a CC BY-SA 3.0 license and is available at Wikimedia Commons.

Breakdown of expert quotes on police shootings (Table 3), © Springer, is from a journal article. *Source*: Crichlow, Vaughn J., and Christopher Fulcher. 2017. "Black Men Down: An Assessment of Experts' Quotes on Deadly Encounters with Police." *Race and Social Problems* 9(3):171–80. doi: 10.1007/s12552-017-9197-x.

Mural featuring Colin Kaepernick on Austin's Fourth Street, by Lars Plougmann, is licensed under a CC BY-SA 2.0 license and is available at Flickr.

Excerpt from coding instructions, © Springer, is from a journal article. *Source:* Neuendorf, Kimberly A., Thomas D. Gore, Amy Dalessandro, Patricie Janstova, and Sharon Snyder-Suhy. 2010. "Shaken and Stirred: A Content Analysis of Women's Portrayals in James Bond Films." *Sex Roles* 62(11):747–61. doi: 10.1007/s11199-009-9644-2.

Table of inter-coder reliability results (Table 2), © Springer, is from a journal article. *Source*: Neuendorf, Kimberly A.,

Thomas D. Gore, Amy Dalessandro, Patricie Janstova, and Sharon Snyder-Suhy. 2010. "Shaken and Stirred: A Content Analysis of Women's Portrayals in James Bond Films." *Sex Roles* 62(11):747–61. doi: 10.1007/s11199-009-9644-2.

Table of frequencies of qualitative codes in student comments (Table 3), © Springer, is from a journal article. *Source*: Chambers, Tony, and Ching-Hsiao Chiang. 2012. "Understanding Undergraduate Students' Experience: A Content Analysis Using NSSE Open-Ended Comments as an Example." *Quality and Quantity* 46(4):1113–23. doi: 10.1007/s11135-011-9549-3.

Graph illustrating views about marriage norms in U.S. women's magazines over the twentieth century (Figure 1), © Sage, is from a journal article. *Source*: Cancian, Francesca M., and Steven L. Gordon. 1988. "Changing Emotion Norms in Marriage: Love and Anger in U.S. Women's Magazines since 1900." *Gender and Society* 2(3):308–42. doi: 10.1177/089124388002003006.

Bar graph illustrating the relationship between gender and scouting activity type (Figure 1), © Sage, is from a journal article. *Source*: Denny, Kathleen E. 2011. "Gender in Context, Content, and Approach: Comparing Gender Messages in Girl Scout and Boy Scout Handbooks." *Gender and Society* 25(1):27–47. doi: 10.1177/0891243210390517.

Crosstabulation of the relationship between gender category and color palette of toys sold on the Disney Store website (Table 2), © Springer, is from a journal article. *Source*: Auster, Carol J., and Claire S. Mansbach. 2012. "The Gender Marketing of Toys: An Analysis of Color and Type of Toy on the Disney Store Website." *Sex Roles* 67(7):375–88. doi: 10.1007/s11199-012-0177-8.

Page from Anne Lister's diary (1832) is licensed under a CC BY-NC-SA 3.0 license and is available from the Calderdale Museums via My Learning.

A. Presenting, Writing, and Publishing

This chapter includes original material, as well as material taken *Principles of Sociological Inquiry: Qualitative and Quantitative Methods* (2012), by Amy Blackstone, which is licensed under a CC BY-NC-SA 4.0 license.

Images from beginning of chapter:

Gold iPhone 7 on top of a book, by Jessica Lewis/thepaintedsquare, is licensed under a Pexels license. The adapted image, by Bizhan Khodabandeh, is licensed under a CC BY-NC-SA 4.0 license. Modifications included cropping and recoloring.

A.1. Deciding What to Share and with Whom to Share It

Woman writing in her notebook, by RF._.studio, is licensed under a Pexels license.

Book cover is from *The War on Normal People: The Truth About America's Disappearing Jobs and Why Universal Basic Income Is Our Future* (2018), by Andrew Yang, © Hachette.

Screenshot of an *Atlantic* article, "Living in an Extreme Meritocracy Is Exhausting" (October 26, 2016), by Victor Tan Chen, © Atlantic Media.

Screenshot of a *New York Times* article, "Who Killed the Knapp Family?" (January 9. 2020), by Nicholas Kristof and Sheryl WuDunn, © New York Times Company.

Screenshot of an *Atlantic* article, "The Spiritual Crisis of the Modern Economy" (December 21, 2016), by Victor Tan Chen, © Atlantic Media.

A.2. Presenting Your Research

Katherine K. Chen presenting at the 2022 SASE conference in Amsterdam, by Victor Tan Chen, is licensed under a CC BY-NC-SA 4.0 license.

People sitting around a round table is licensed under a CC BY-NC-ND 4.0 license and is available from the Gender Spectrum Collection via Vice Media.

Logos in the collage are © American Sociological Association, Midwest Sociological Society, Pacific Sociological Association, Rural Sociological Society, Society for the Advancement of Socio-Economics, and Southern Sociological Society.

A.3. Writing Up Your Work for Other Scholars

Featured pages, © Sage, are from a journal article. *Source*: Davis, Daniel, and Amy Binder. 2019. "Industry, Firm, Job Title: The Layered Nature of Early-Career Advantage for Graduates of Elite Private Universities." *Socius* 5:2378023119859711. doi: 10.1177/2378023119859711.

Presentation of hypotheses, © Sage, is from a journal article. *Source*: Schneider, Daniel, Orestes P. Hastings, and Joe LaBriola. 2018. "Income Inequality and Class Divides in Parental Investments." *American Sociological Review* 83(3):475–507. doi: 10.1177/0003122418772034.

Reference list, © Sage, is from a journal article. *Source*: South, Scott J., and Lei Lei. 2021. "Why Are Fewer Young Adults Having Casual Sex?" *Socius* 7:2378023121996854. doi: 10.1177/2378023121996854.

Photo of Stephen Ambrose, by Jim Wallace, is licensed under a CC BY 2.0 license and is available from the Smithsonian Institution via Flickr.

Photo of Doris Kearns Goodwin, by Rhododendrites, is public domain and is available at Wikimedia Commons.

A.4. Reaching Your Intended Audience

People at a concert, by Vishnu R Nair, is licensed under a Pexels license.

The White House, by Aaron Kittredge, is licensed under a Pexels license.

Screenshot of the *Atlantic* article "The Lonely Poverty of America's White Working Class" (January 16, 2016), by Victor Tan Chen, © Atlantic Media.

Contributors

Victor Tan Chen is an associate professor of sociology at Virginia Commonwealth University (VCU) who studies economic inequality and labor markets. He has authored three books: *The Missing Class: Portraits of the Near Poor in America* (with Katherine S. Newman), named a *Library Journal* best business book of the year; *Cut Loose: Jobless and Hopeless in an Unfair Economy*, for which he received the John T. Dunlop Outstanding Scholar Award; and *Organizational Imaginaries: Tempering Capitalism and Tending to Communities through Cooperatives and Collectivist Democracy* (with Katherine K. Chen), which received the Joyce Rothschild Book Prize. Chen's work has been featured in the *Atlantic*, *New York Times*, BBC News, *Fortune*, and NPR. He is also the editor in chief of *In The Fray*, an award-winning magazine devoted to personal stories on global issues. You can read more about his work at victortanchen.com.

Gabriela León-Pérez is an assistant professor of sociology at VCU. Her scholarly interests lie at the intersection of the sociology of migration, Latino sociology, and medical sociology. Specifically, Gabriela's research explores the determinants of migration from Latin America to the United States, as well as how social and contextual factors shape the health and integration of Latino immigrants and their children. At VCU, she teaches courses on research methods, immigration, and racial and ethnic health disparities. She also provides research mentorship to undergraduate and graduate students through her Migration and Health Equity Lab. You can read more about her work at gleonperez.weebly.com.

Julie Honnold is a professor emerita of sociology at VCU. She taught all of the undergraduate and graduate research methods courses offered by the department numerous times over several decades. Her primary research activities have been conducted as a methodologist and data analyst with the VCU Survey and Evaluation Research Laboratory (SERL). Since the early 1980s, she has been involved in a wide variety of projects for local, state, and national clients, using a wide variety of quantitative and qualitative research methodologies. Her most extensive experience at SERL has been on policy-related projects in the areas of sexuality, youth risk behavior, and HIV/AIDS. In addition, she frequently works as a data analyst on joint projects.

Volkan Aytar is a teaching faculty member at VCU Sociology and a Fulbright scholar. His PhD degree is from the University of Amsterdam in the Netherlands. He has studied gentrification and leisure consumption in multicultural cities, such as Istanbul and Amsterdam. He teaches classes on research methods, the sociology of food, and media and society, as well as introduction to sociology and theory courses. Previously, he was the director of the Creative Industries Center at Bahcesehir University (BAU) in Turkey, and he served as the chair of the university's Department of New Media. He was also a guest researcher at the Max Planck Institute for Ethnic and Religious Diversity in Goettingen, Germany. You can read more about his work at volkanaytar.academia.edu.

Visual Design

Bizhan Khodabandeh is a visual communicator who moves freely across professional boundaries as a designer, illustrator, artist, and cartoonist. Khodabandeh is particularly fascinated by how art and design can be a catalyst for social change. He has received recognition for his work as both an illustrator and designer from various institutions, such as the Association for Education in Journalism and Mass Communication, the Society of Illustrators, the American Institute of Graphic Arts, Creativity International, the International Design Awards, *Adbusters*, and *Creative Quarterly*.

Khodabandeh is currently an assistant professor at VCU's Robertson School of Media and Culture and freelances under the name Mended Arrow. You can read more about his work at mendedarrow.com.

Technology

Tom Woodward is the director of learning spaces and technologies at Middlebury College. He has been working at the intersection of education and technology for more than 20 years.

Copyright and Open Licensing

Jessica Kirschner is the digital publishing coordinator at VIVA, where she manages the open educational resources (OER) publishing efforts of Virginia's academic library consortium. Jessica began her career working in the acquisitions department at SUNY Press. Since transitioning to librarianship, she has held various positions focusing on publishing and open education—including, most recently, serving as the OER librarian at VCU.