

Renita
Coleman



For Review-No Commercial Use(2023)

Designing
Experiments
for the Social Sciences



Designing Experiments for the Social Sciences

For Review-No Commercial Use(2023)

For Review-No Commercial Use(2023)

Sara Miller McCune founded SAGE Publishing in 1965 to support the dissemination of usable knowledge and educate a global community. SAGE publishes more than 1000 journals and over 800 new books each year, spanning a wide range of subject areas. Our growing selection of library products includes archives, data, case studies and video. SAGE remains majority owned by our founder and after her lifetime will become owned by a charitable trust that secures the company's continued independence.

Los Angeles | London | New Delhi | Singapore | Washington DC | Melbourne

Designing Experiments for the Social Sciences

How to Plan, Create, and Execute
Research Using Experiments

For Review-No Commercial Use(2023)

Renita Coleman

University of Texas at Austin



Los Angeles | London | New Delhi
Singapore | Washington DC | Melbourne



FOR INFORMATION:

SAGE Publications, Inc.
2455 Teller Road
Thousand Oaks, California 91320
E-mail: order@sagepub.com

SAGE Publications Ltd.
1 Oliver's Yard
55 City Road
London EC1Y 1SP
United Kingdom

SAGE Publications India Pvt. Ltd.
B 1/I 1 Mohan Cooperative Industrial Area
Mathura Road, New Delhi 110 044
India

SAGE Publications Asia-Pacific Pte. Ltd.
3 Church Street
#10-04 Samsung Hub
Singapore 049483

Acquisitions Editor: Helen Salmon
Development Editor: Chelsea Neve
Editorial Assistant: Megan O'Heffernan
Production Editor: Olivia Weber-Stenis
Copy Editor: Diane Wainwright
Typesetter: C&M Digitals Ltd.
Proofreader: Jennifer Grubba
Indexer: Jean Casalegno
Cover Designer: Ginkhan Siam
Marketing Manager: Susannah Goldes

Copyright © 2019 by SAGE Publications, Inc.

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

Printed in the United States of America

ISBN 978-1-5063-7732-2

For Review-No Commercial Use(2023)

This book is printed on acid-free paper.

18 19 20 21 22 10 9 8 7 6 5 4 3 2 1

PRAISE FOR THIS BOOK

“This book offers many examples and hands-on instruction that can help students learn how to conduct experiments.”

—Francis O. Adeola, University of New Orleans

“This book is a must for learning about the experimental design—from forming a research question to interpreting the results, this text covers it all.”

—Sarah El Sayed, University of Texas at Arlington

“This text provides students with an excellent explanation of experimental methodology: not just descriptions of elements of the experiment method, but understanding of where those elements came from, how and why they work together, and how students can become scholars.”

For Review-No Commercial Use(2023) —Kim L. Heppner, Pasadena City University

“This is a comprehensive text on experiments, clearly and engagingly written, with many excellent examples. Renita Coleman has included everything the student will need to develop, design, and execute a methodologically sound experiment.”

—Glenn Leshner, University of Oklahoma

“This textbook is detailed and well organized, with each topic progressing logically to build on foundational knowledge. Students will find it easy to master research concepts and practice using the exercises and examples provided. The ‘Study Spotlight’ scenarios and practical examples are effective tools to demonstrate research designs, and the ‘Test Your Knowledge’ quizzes reinforce student learning.”

—Janet Reid-Hector, Rutgers University

“This book is well-written and readable. If you need to grasp concepts of experimental designs, this book guides you as to what the experimental designs are, and what you should keep in mind in conducting your studies.”

—Ji Hoon Ryoo, University of Virginia

For Review-No Commercial Use(2023)

BRIEF CONTENTS

Preface	xv
About the Author	xxi
Chapter 1 • Discovering Cause and Effect	1
Chapter 2 • Ethics and Famous Experiments in History	23
Chapter 3 • Theory, Literature, and Hypotheses	55
Chapter 4 • Types of Experiments	89
Chapter 5 • Internal and External Validity	111
Chapter 6 • Factorial Designs	145
Chapter 7 • Random Assignment	173
Chapter 8 • Sampling and Effects Size	211
Chapter 9 • Stimuli and Manipulation Checks	249
Chapter 10 • Instruments and Measurement	289
Chapter 11 • The Institutional Review Board and Conducting Ethical Experiments	337
Glossary	369
Index	377

For Review-No Commercial Use(2023)

DETAILED CONTENTS

Preface	xv
Acknowledgments	xix
About the Author	xxi
1. Discovering Cause and Effect	1
Causation	2
Experiments Compared to Other Methods	3
Basic Criteria for Experiments	4
Elements of Experiments	6
Variation	6
Confounds	7
Control Groups	9
Assignment	9
Starting a Study of Your Own	11
Writing a Statement of the Problem	12
Answering the “So What” Question	14
2. Ethics and Famous Experiments in History	23
The Scurvy Studies	24
The Contributions of Charles Peirce	27
Ronald Fisher’s Plots and Tea	28
B. F. Skinner: Small Samples, High Tech	30
Stanley Milgram Shocks the World	33
Philip Zimbardo: Raising Consciences in a Stanford Basement	40
Conclusion	44
3. Theory, Literature, and Hypotheses	55
The Literature Review	58
Tips on Writing the Literature Review for Experiments	60
<i>It Is Not a Book Report</i>	60
<i>Do Not Be Wikipedia</i>	61
<i>Make a Theoretical Contribution</i>	63
<i>Connect the Dots</i>	63

Conceptual Definitions vs. Operationalizations	65
Literature Reviews With Multiple Experiments	66
Hypotheses and Research Questions	68
Hypothesis Basics	72
<i>Null vs. Alternative</i>	72
<i>Difference vs. Direction</i>	73
Hypothesis Writing Formula	75
Hypotheses With More Than One IV	77
Research Questions	81
4. Types of Experiments	89
Campbell and Stanley's Typology of Experiments	90
Three Pre-Experimental Designs	91
<i>The One-Shot Case Study</i>	91
<i>One Group Pretest-Posttest Design</i>	91
<i>Static Group Comparison</i>	93
Three True Experimental Designs	93
<i>Pretest-Posttest Control Group</i>	93
<i>Solomon Four-Group Design</i>	94
<i>Posttest-Only Control Group Design</i>	94
Quasi Experiments	96
Natural Experiments	100
Field Experiments	102
5. Internal and External Validity	111
Ecological and External Validity	112
Generalizability	113
Random Sampling	116
<i>Two-Step Randomization</i>	117
<i>Nonresponse Bias</i>	117
Representativeness	118
Cause and Effect	122
Logical Inference	123
Replication	124
Self-Replication	124
Exact vs. Conceptual Replication	125
Multiple Experiments	126
External Replication	127
Triangulation	129
Internal Validity	129
Three Basic Criteria	130
Random Assignment	131

For Review-No Commercial Use(2023)

Confounds	131
<i>History</i>	132
<i>Maturation</i>	133
<i>Testing</i>	133
<i>Instrumentation</i>	134
<i>Statistical Regression</i>	135
<i>Selection</i>	135
<i>Attrition</i>	135
 6. Factorial Designs	 145
Single-Factor Designs	146
Factorial Designs	146
Main Effects and Interactions	147
Factorial Notation	150
Design Tables	154
Choosing a Design	156
How Subjects Are Used in Designs	158
Between-Subjects Designs	158
Within-Subjects Designs	159
Mixed Factorial Designs	161
Incomplete Factorials	162
Control Groups	163
The No-Treatment Control Group	165
Creativity in Control Groups	166
The Status Quo Control Group	166
No Control Group	167
 7. Random Assignment	 173
The Purpose of Random Assignment	174
Avoiding Confounds	175
What Is Random?	176
Operationalizing Random Assignment	178
Computerized Randomization	178
Survey Experiments	182
In-Person Randomizing	183
Reporting Random Assignment	183
Balanced and Unbalanced Designs	183
Checking That Random Assignment Was Effective	184
Aggregate Level Random Assignment	185
Reporting Random Assignment Results	185
When Random Assignment Fails	187

Blocking, Matching, and Other Strategies	189
Blocking vs. Simple Random Assignment	190
Stratified Random Assignment	191
Random Assignment of Other Things	191
Counterbalancing	194
Latin Squares	194
Random Assignment Resistance	196
8. Sampling and Effect Sizes	211
Student Samples	212
Higher Than Average Cognitive Skills	213
Compliance With Authority	217
Weaker Sense of Self	217
Myriad Other Differences	217
Homogeneity	218
Basic Psychological Processes	218
Guarding Against Bias With Students	219
Amazon's Mechanical Turk	221
Is MTurk Representative?	222
Advantages and Drawbacks	222
Other Subject Sources	225
Recruiting	226
Incentives	228
Sample Size and Power	229
Effect Sizes	231
Power Analyses	234
<i>Observed Power</i>	237
9. Stimuli and Manipulation Checks	249
Examples of Stimuli	251
Advice on Creating Stimuli	257
Keep It Real	257
Control as Much as Possible	258
Vary Only the Things Being Studied	259
Maximize Comparisons	260
Employ Message Variance	261
<i>How Many Stimuli to Create?</i>	263
<i>Fixed vs. Random Factors</i>	263
Manipulation Checks	264
Manipulation Check, Pretest, or Pilot Study?	265
Direct vs. Indirect Manipulations	266
Mediators and Psychological States	266

For Review-No Commercial Use(2023)

Manipulation Checks in Survey Experiments	271
Practical Issues	272
Reporting the Stimuli and Manipulation Checks	274
Stimuli Write-Up	274
Manipulation Check Write-Up	276
 10. Instruments and Measures	 289
Instruments	290
Questionnaires	290
<i>Single Items vs. Indexes</i>	292
<i>Pros and Cons</i>	293
Unobtrusive and Nonreactive Measures	297
Technological Instruments	297
<i>Functional Magnetic Resonance Imaging</i>	299
<i>Heart Rate</i>	301
<i>Skin Conductance</i>	301
<i>Other Physiological Instruments</i>	301
<i>Eye Tracking</i>	302
<i>Eyes on Screen</i>	303
<i>Web Tracking</i>	303
<i>Catback Response</i>	304
<i>Secondary Reaction Tasks</i>	305
<i>Continuous Response Instruments</i>	306
<i>Thought Listing and Think Aloud</i>	307
Observations	309
<i>Interobserver Reliability</i>	310
Other Unobtrusive Instruments	312
Unobtrusive Questions	312
Measurement Issues	312
Levels of Measure	315
Response Choices	315
<i>Ceiling and Floor Effects</i>	316
Construct Validity	317
 11. The Institutional Review Board and Conducting	 337
Ethical Experiments	337
Institutional Review Boards	338
Ethical Issues in Experiments	342
Deception	342
Harm	344
Abuse of Power	345
Random Assignment	346

For Review-No Commercial Use(2023)

Protecting Subjects	346
Informed Consent	347
Privacy and Confidentiality	348
Researcher Issues	349
Data Ethics	349
Plagiarism and Self-Plagiarism	351
Authorship Issues	352
Pilot Studies	354
Glossary	369
Index	377

For Review-No Commercial Use(2023)

PREFACE

“There are excellent books and courses of instruction dealing with the statistical manipulation of experimental data, but there is little help to be found on the methods of securing adequate and proper data to which to apply statistical procedure.”¹

Forty years after W. A. McCall wrote this statement, Don Campbell and Julian Stanley said, “This sentence remains true enough today.”²

More than fifty years have passed, and I say this *still* remains true today, especially for the social sciences.

This book is about securing adequate and proper data.

It is different from most other experimental design books in two ways: First, it concentrates on the methodological and design issues of planning an experiment rather than on analyzing data with various statistics after the data are collected. Careful design of studies from the beginning is the key to good research. That is nowhere more the case than with experiments. This book is about how to effectively *design* experiments rather than how to *analyze* them. It focuses on the stage where researchers are making decisions about procedural aspects of the experiment, before interventions and treatments are given.

It will help readers learn how to plan and execute experiments from the beginning by walking step-by-step through deciding whether to use a single-factor or factorial design, how to assign subjects to groups, choosing and collecting a sample, creating the stimuli and instrument, doing a manipulation check, applying for approval from the Institutional Review Board (IRB), and doing a pilot study, as well as other choices along the way. It gives guidelines for deciding which elements are best used in the creation of a particular experiment and things to consider when making the inevitable trade-offs. It is practical and applied.

Designing Experiments for the Social Sciences only briefly covers the statistics aspect of experiments, for they are inextricably linked. However, the focus is mainly on enabling readers to learn how to design experiments from the beginning. The inner workings of the statistics, formulas, and how to calculate them by hand or using software is not part of this book. In-depth knowledge of statistics is not required to understand the material. This book is confined to helping readers understand which statistics are appropriate for

what kinds of analysis and level of measurement, but it is very basic. I recommend taking a traditional experimental design class, statistics courses, or reading such a book that focuses on statistics in addition to this one.

The second way in which this book is different from others is that it is aimed at social scientists broadly. Readers will find examples beyond psychology ranging from political science, business, economics, information sciences, social work, education, sociology, health fields, advertising, and more. My discipline is journalism and the larger field of communication. My first experimental design class had students enrolled from journalism, advertising, public relations, radio, television, film, political, and interpersonal communication. Experiments are becoming more prevalent throughout the social sciences as researchers discover the importance of understanding cause and effect to their discipline. With the introduction of easy-to-use software (such as SPSS, G*Power, Qualtrics) and greater access to human subjects (such as Mechanical Turk), experimental designs are becoming more prevalent. This book speaks directly to these researchers with examples from their own worlds. It also explores some of the challenges facing specific disciplines—for example, objections to random assignment common in education, reporting of response rates with samples in political science, and the use of observations or performance-based measures rather than self-reports in economics.

For Review-No Commercial Use(2023)

Many researchers in the social sciences learn by working with another experimentalist one-on-one. This book is written to be a supplement to those learning this way, a guidebook for those working on their own, or a textbook for an organized class. It came about when I first taught an experimental methods class at the University of Texas and could not find a book that did exactly what I wanted. That course is the foundation for this book. I envisioned it as a core textbook for stand-alone experimental design methods courses. It will also be a good supplementary text for general research methods courses in the social sciences where experiments are prominent—for example, advertising, political science, health sciences, and others.

I wrote *Designing Experiments for the Social Sciences* with graduate students in mind, although my fifteen years as a professional journalist have taught me how to write for a general audience. This book is written in easy-to-understand language so that undergraduate students can understand it too. It is not appropriate for the totally uninitiated, however; it is more suited to an educated novice, someone who has conducted research previously but has not done an experiment, or has assisted with one but not conducted one of his or her own. It presumes at least one general overview course in the scientific method. Those who have some research experience but have never conducted an experiment before will not find it too challenging. For those learning from the benefit of mentors rather than an organized course, this book is meant as a reference and

supplement to those excellent instructors. It also can help generate ideas for tools and strategies that one's mentors may not routinely include in their repertoire. Other readers may include academics with experience in other research methods that want to expand their abilities to explore cause and effect. In addition, professionals working in a research capacity, whether for a political candidate, advertising agency, educational institution, or other industry, will find this book helpful. I hope it inspires social science departments without a stand-alone experiments course to create one.

The chapters contain a mix of conceptual and theoretical issues as well as practical hands-on advice. It takes a step-by-step approach, walking students through writing hypotheses and research questions, designing stimuli, and writing up the methods sections. Many examples come from my own experience, allowing me to provide the backstory about why certain procedures were chosen over others, how the study evolved, and tips for responding to reviewer comments. These things frequently do not make it into published papers. I also draw from some of the excellent experiments conducted by others in various social science disciplines, although I am unable to offer inside knowledge about the choices made.

Although the chapters are presented roughly in the order in which designing an experiment might proceed, some instructors or readers may wish to see 2 in the order presented. For example, after reading about deception in chapter 2, some readers may wish to skip ahead to chapter 11 where broader ethical issues are discussed in conjunction with the IRB. To make reading out of order easier, there is a Glossary at the end of the book containing each chapter's key terms, which are bolded and defined in the text. Other pedagogical tools include three types of boxed features:

- “More About . . .” boxes expand on some issue in the chapter for those who would like more detail.
- “Study Spotlight” boxes highlight a particular experiment for how it handles the topics in the chapter.
- “How To Do It” boxes give instructions and examples for executing the steps in the chapter and writing it up for a paper.

In addition, there are three other features:

- “Common Mistakes” is a short bulleted list of errors novices are especially prone to.
- “Test Your Understanding” questions at the end of each chapter is a short multiple-choice quiz so you can see how well you understood the material.

- “Application Exercises” are instructions for longer hands-on assignments that let you apply what you have learned, including a one-step-at-a-time approach for creating a research proposal for your own experiment. This is broken down into manageable pieces so that the task does not seem so daunting.

Each chapter also contains a list of suggested readings for more depth on the topics. A defining feature of every section is the practical advice and examples on everything from how to describe the stimuli and instrument to working with coauthors, submitting to journals, and responding to common questions from reviewers. There are URLs for free software to use when creating Twitter feeds or Facebook posts, online randomizers, and tutorials for conducting power analyses. Some topics are not unique to experiments, but I have tried to hone in on particular issues with this method. For example, chapter 3 explains the role of theory in experimentation, whose strength is in testing and expanding theory. It also covers how empirical evidence from the literature should be used to build a case for the hypotheses and theoretical processes being tested, and offers a “formula” for hypothesis writing that resembles diagramming a sentence.

Ethics is first introduced in chapter 2, following naturally the stories about Stanley Milgram’s obedience to authority experiments and Philip Zimbardo’s prison experiment, both of which raised collective consciences about the importance of ethics in experimental science. A “More About” box here delves into deception, which is more problematic with experiments than other methods. The discussion about ethics is woven throughout each chapter—for example, in chapter 7 on random assignment, some of the objections to it as unethical are discussed. The topic is revisited in greater detail in chapter 11, using IRBs as a framework considering that their standards are intertwined with ethics. Because of the step-by-step approach of this book, the IRB is discussed after the experiment is designed and researchers must apply for approval. However, the importance of ethical experiments is not an afterthought; rather, it is included in chapter 11 because IRBs are the principal means of enforcing ethical behavior, so this chapter seemed the most logical for discussing both.

What there is to know about experiments does not begin and end with this textbook; I encourage the learning process to continue by reading some of the excellent writings in the “Suggested Readings” sections, by reading experiments in one’s own discipline, and by working with others. I hope you enjoy and find it helpful.

For additional teaching and learning resources that go along with this book, please visit the website at study.sagepub.com/coleman.

Instructors will find chapter-by-chapter PowerPoint® slides to assist in lecture preparation, along with a Word® test bank with multiple-choice and open-ended exam questions. Students will find the full-text version of the SAGE journal articles spotlighted throughout the book.

ACKNOWLEDGMENTS

I would like to thank my own mentors, who infected me with the experimental bug when I was a graduate student—Esther Thorson and Glenn Leshner. Thank you for your guidance, patience, and for sharing your passion. For telling others—which got back to me—that “she always does what she says she will,” which helped ensure that happened. For being the role model who measured the distance from screen to chair every time a new subject sat in it, opening my eyes to the importance of detail. And for always taking the time to hear about a new project and give advice. I also thank my longtime partner in research, Denis Wu, whom I was fortunate to meet in my first faculty position and who has stuck with me since. He always seems to be good at that which I am not, and has been my constant ally on our self-guided tour of experiments. I thank all the brilliant former graduate students I have experimented with, including but not limited to Ben Wasike, Lesa Hatley-Major, Rebecca McEntee, Lewis Knight, Carolyn Yaschur, Trent Boulter, Viorela Dan, Angela Lee, Raluca Cozma, Avery Holton, Dani Kilgo, Siobhan Smith, Kate West, Joseph Yoo, and Ji Won Kim, among others. And I thank all those who enrolled in my experimental design course, especially from other departments, opening my eyes to new theories and unique challenges in your fields, and for allowing me to experiment on you in the process of writing this book. I learn as much, if not more, from you as you do from me.

Finally, this book would not have been possible without the time and expertise of all the reviewers who gently pointed out flaws and gaps, steered me toward valuable resources, and patiently helped make this book much better. I also thank Helen Salmon, the senior editor at SAGE, who took a chance on this proposal and suggested numerous important additions; Jeremy Shermak, for creating the website for the book; and the excellent staff at SAGE who shepherded this to completion, including Chelsea Neve, Megan O’Heffernan, Eve Oettinger, Diane Wainwright, and the rest of the team in the Research Methods, Statistics, and Evaluation department.

SAGE would like to thank the following reviewers for their feedback:

Richard E. Adams, Kent State University

Francis O. Adeola, University of New Orleans

Anna Bassi, University of North Carolina

Jacqueline Craven, Delta State University

Sarah A. El Sayed, University of Texas at Arlington

Janet Reid Hector, Rutgers University

Kyle J. Holody, Coastal Carolina University

Glenn Leshner, University of Oklahoma

Ji Hoon Ryoo, University of Virginia

Gerene K. Starratt, Barry University

Michael Teneyck, University of Texas at Arlington

Geoffrey P. R. Wallace, University of Washington

For Review-No Commercial Use(2023)

Notes

1. W. A. McCall, *How to Experiment in Education* (New York: MacMillan, 1923), Preface.
2. D. T. Campbell and J. C. Stanley, *Experimental and Quasi-Experimental Designs for Research* (Chicago: Rand McNally, 1963), 1.

ABOUT THE AUTHOR



Rebecca Scoggin McEntee

Renita Coleman has a bachelor's degree in journalism from the University of Florida and a master's and PhD in journalism from the University of Missouri.

Her research focuses on ethics, framing, and agenda setting, with a special focus on visual communication. She has studied the effects of photographs on ethical reasoning, the framing and attribution of responsibility in health news, and the moral development of journalists and public relations practitioners, among other topics. She has published more than 40 peer-reviewed articles in academic journals including *Journal of Communication*, *Journalism and Mass Communication Quarterly*, *Journal of Broadcasting & Electronic*

Media, *Journal of Mass Media Ethics*, *Journalism*, and *Journalism Studies*. She has coauthored two books: *Image and Emotion in Voter Decisions: The Affect Agenda* in 2015, and *The Moral Media: How Journalists Reason About Ethics* in 2005.

Before beginning her academic career, Coleman was a journalist at newspapers and magazines for fifteen years. She was a reporter, editor, and designer at the Raleigh, North Carolina, *News & Observer*, the Sarasota, Florida, *Herald-Tribune*, and the Orlando, Florida, *Sentinel*, among other news organizations.

Coleman teaches undergraduate and graduate courses in ethics, lifestyle journalism, and experimental design.



For Review-No Commercial Use(2023)



DISCOVERING CAUSE AND EFFECT

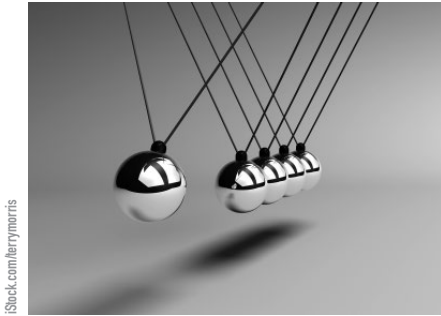
Life is a perpetual instruction in cause and effect.

—Ralph Waldo Emerson

LEARNING OBJECTIVES

- Explain how cause and effect work in an experiment.
- Compare the benefits of experiments to other methods.
- Identify the three basic criteria of experiments.
- Describe the elements of variation, confounds, control groups, and assignment.
- Develop a statement of the problem and answer the “so what” question for a study of your own.

This book is about experiments, the scope of which varies greatly. An experiment is a scientific test of some hypothesis or principle carried out under carefully controlled conditions in order to determine or discover something unknown. Experiments provide insights about the relationship between things where changes in one thing cause something to happen to another. We have all done informal experiments in everyday life without even knowing it, as the opening quote by Ralph Waldo Emerson illustrates. For example, if you change the amount of a certain ingredient in a recipe, does it taste better? If you drink lower-calorie beer, do you lose weight? As long as you only



change one ingredient at a time, or do not exercise more or eat lower-calorie everything, you probably assumed it was that one ingredient or that lower-calorie beer that caused the difference in taste or your weight loss.

CAUSATION

These examples of everyday experiments illustrate the concept of cause and effect. An “effect” is what happened. Better-tasting lasagna or weight loss are the effects in the earlier examples. The “cause” is the explanation for why these things happened—more garlic in the lasagna made it taste better, the lower-calorie beer helped you lose weight. We find the word *cause* used in everyday language, such as “the cause of death” or “the cause of an accident.” The meaning is no different in experiments, but you will also see the terms *causality* or *causation* used.

Of course, for the purposes of this book, we are more interested in systematic experiments than the simple, everyday ones that are described. In medicine, these are called *clinical trials* or *randomized clinical trials* (RCTs; for more about RCTs, see More About box 1.1). In web design and market research, they may also be called *A/B testing*. The language is a little different from experiments in the social sciences, but the goal is the same—to discover what treatment (or cause) works best on a particular problem. In medicine, the problem is an illness or disease, the effect being a cure or improvement. In social science, the problems for which we are seeking solutions can be TV commercials that promote brand awareness, strategies for teaching students with Down syndrome, or interventions that help accountants be more honest.

To do this, social scientists use experiments as “the basic language of proof.”⁶ Experiments give us evidence of cause and effect by demonstrating what happens when something is changed while everything else remains the same. In this way, we have more assurance that the thing that changed is responsible for causing the outcome or effect we have observed.

Experiments are the most common kind of research conducted in the medical field, but in social science, they can be among the least often used. In communication journals, around 12% of studies use experimental designs.⁷ In international relations, it can be as low as 4%.⁸ In special education, experiments are promoted as the answer to calls for increased quality and rigor in an evidence-based profession.⁹ And they are growing steadily in political science,¹⁰ among other fields. As the benefits of this method become known throughout the social sciences, software to analyze data gets easier to use, and as technology makes subjects cheaper and easier to recruit, the use of experiments should

MORE ABOUT . . . BOX 1.1

Randomized Clinical Trials

RCTs are basically the same things as true experiments, where subjects are randomly assigned (the R) to either a treatment or a control group and given one or more experimental procedures or drugs. They are also referred to as randomized *controlled* trials, with language purists using controlled for studies that include a **control group** where subjects receive a placebo or no treatment. When *clinical* is used, it may or may not have a control group. Presumably, the clinical usage came into being because these studies were conducted on medical patients or others in a clinical setting. The term *trial* is used because the treatment or drug being studied is being tried out—that is, it is not approved for widespread use, and the study will determine if it is safe and effective.¹

RCTs are considered the gold standard for medical studies, just as their counterpart is in social science—the true or lab experiment.

The history of RCTs dates back to 600 BC when Daniel of Judah compared the royal Babylonian diet to a vegetarian diet.² Others credit James Lind, who conducted the scurvy experiments in 1747 described in chapter 2.³ The first modern-day RCT is usually recognized as the test of streptomycin's effects on tuberculosis in 1948.⁴ The study begins by explaining how the preponderance of inadequately controlled clinical trials on tuberculosis had led to “exaggerated claims” about gold as a treatment. The study presents a “full description” of the methods because of the difficulty of planning such a rigorous trial, so that others may reproduce it.⁵

For Review-No Commercial Use(2023)

only increase. Thus, knowing how to do them well is all the more important. Already, many articles discuss the increased use of experiments for both academics and professionals in disciplines such as political science¹¹ and information systems.¹² Others discuss the growing importance of creating organized courses in experimental methods.¹³ Some go so far as to say experiments are the most important method in their discipline.¹⁴ Experiments are central to fields that use evidence-based practice such as social work¹⁵ and education, where the randomized clinical trial is the gold standard.¹⁶

EXPERIMENTS COMPARED TO OTHER METHODS

The main benefit of the experimental method is that it offers a powerful tool to discover causation. Many other research methods, such as surveys, can identify **correlations**, or relationships, that vary together and are unlikely to have occurred by chance. This is not

necessarily the same thing as causation. Experiments, by contrast, can provide insights into how changing one thing leads to changes in another. Deliberately and systematically **varying**, or changing, something allows us to see a potential causal agent.¹⁷ Following up correlational studies with experiments is a good way to give us more confidence that the relationships we find in these studies are actually causal. **Triangulation**, or the use of different methods to study the same phenomenon, is important to scientific inquiry because it helps give us confidence that what we see using one method can be **replicated**, or reproduced, using another. Of much concern is the overreliance on surveys and observational methods that show a correlation but with no follow-up using experiments that find evidence of a causal relationship.¹⁸

One of the most famous illustrations of the problem with inferring causation from correlation is the long-ago pronouncement that storks brought babies. A study done in Copenhagen in the 1930s documented that the years with larger stork populations also saw more babies born—a high correlation of .85.¹⁹ But just because these two variables are

highly correlated does not mean one caused the other. Instead, there were **plausible alternative explanations**, or other possible causes, that were not studied. This was right after World War I, and all those soldiers returning home after so long led to more babies being born. In addition, people were migrating from the country to the city where the jobs were, so more people to have babies equaled more babies. As the population increased, more houses were built, which led to more places for storks to nest, leading to more storks.²⁰ The cycle continued.



For Review-No Commercial Use(2023)

BASIC CRITERIA FOR EXPERIMENTS

This example illustrates the three basic features that all experiments must have; that is, the cause must precede the effect—in this case, storks did come before the babies—but the cause must also be related to the effect, which it was not. There was no logical or theoretical reason and no empirical evidence to suggest that storks were related to babies. This is one reason why experimentalists do not test mere hunches but instead find linkages in the form of theoretical, logical, or existing evidence to test. This helps ensure that the cause is actually related to the effect. In the storks and babies example, the third feature of an experiment also was missing—there must be no other plausible alternative explanation for the effect. A real experiment must contain all three: Cause must come before the effect, be related to it, and there must not be any other plausible alternative explanations. If these three conditions are met, experiments give us a powerful way to have more

STUDY SPOTLIGHT 1.2

Discovering Effects and Explaining Why



SAGE Journal Article:
study.sagepub.com/
coleman

Hay, Carter, Xia Wang, Emily Ciaravolo, and Ryan C. Meldrum. 2015. "Inside the Black Box: Identifying the Variables That Mediate the Effects of an Experimental Intervention for Adolescents." *Crime & Delinquency* 61 (2): 243–270.

This study is a good example of an experiment that not only finds effects but also explains what caused them. The authors start by reconfirming that the treatment by a particular program actually did reduce juveniles' risk for delinquency. Then they examined the mediating variables that intervene, or come between, participation in the program and reduced juvenile delinquency. They examined a total of eleven risk factors that had been suggested as the reasons why the program worked, which researchers call *causal mechanisms*. They say, "In short, if we lack insight on the precise mechanisms by which a program reduces delinquency, then efforts to build on its strengths, replicate it elsewhere, and use it to inform public policy are necessarily hindered" (p. 248).

Out of the eleven possible variables that could have explained why the program worked, the researchers found only one that was significant: "Reduced association with peers who engaged in deviance and pressured them to do so as well" (p. 263). In other words, hanging out with a bad crowd.

They explain the importance of this in terms of the time and money spent by such programs pursuing these ten other variables that, it turns out, did not actually make a difference. Put in nicer terms than the wasting of time and money, the researchers say, "Our analysis—a rare test that has considered mediating variables—suggests that many of the risk factors targeted by these programs may be unresponsive to program services" (p. 264).

confidence that one thing led to, or caused, another, not just that there is an association that was unlikely to have occurred by chance.

While being able to test cause and effect is a powerful tool, it is still merely a description; it does not explain why something occurred. For example, in the field of social work, many studies have identified programs that successfully reduce juvenile delinquency, but few have undertaken an examination of why they are effective.²¹ Hay and colleagues noticed the gap and designed an experiment to discover the mechanisms that explain why participating in a program reduced delinquency rates (for more on this study, see Study Spotlight 1.2). In my own work, I hypothesized that seeing photographs would cause journalists to use better-quality ethical reasoning when making news decisions. I found the effect I was looking for,²² but that was a description, not an explanation. That finding alone did not say anything about why it happened. To provide a causal explanation, experiments need to build in potential mechanisms that explain these effects. If photographs do improve ethical reasoning, it is important to know why. **Mediators** and **moderators** such as this will be discussed in more detail in another chapter. Suffice it

to say here that good experiments should also include mechanisms for explaining why a certain effect occurred.

With this basic discussion of what experiments can do compared to other methods, next we turn to some specific elements of experiments. All of these will be elaborated in more detail in later chapters but are introduced here to provide some basic fundamental understanding of experiments.

ELEMENTS OF EXPERIMENTS

Variation

Varying or changing things is the first key to a good experiment. Obviously, if nothing varies, then there is nothing to study. Something has to change. In experiments, variation is achieved by **manipulating** the **independent variable**, or IV—the thing researchers think will cause a change. This is also called the *manipulation*, *treatment*, or *intervention*. The researcher should carefully control these. For example, many studies use professional actors to portray a political candidate rather than real politicians who may bias subjects' responses.²³ In one study, the actor was tasked with displaying positive nonverbal behaviors in one interview, negative ones in another, and neutral body language in a third. The study authors researched exactly what those looked like—for example, crossed arms was negative, leaning forward and looking the interviewer in the eye was positive—and made sure the actor displayed those behaviors. It is not enough to merely ask an actor to behave positively or negatively but for the researchers to know exactly what that should look like and ensure that it is properly demonstrated.

Another key to good variation is that only one thing at a time can vary; otherwise, it is impossible to tell which of the several things that varied caused the outcome. Actually, many experiments do vary more than one thing at a time (more on that later), but the key is they do not allow those things to **covary**—that is, they cannot vary together. For example, one researcher noticed that studies showing negative political advertising were more powerful than positive advertising and had allowed the tone of the ads to covary with the amount of information in them.²⁴ There was always more information in the negative ads than in the positive ads; therefore, the more powerful effects could be due to more information, not necessarily to its negative tone. This provided a plausible alternative explanation for the effects found in these other studies. To determine if this was the case, the researcher did a study that held the amount of information **constant**—that is, there was the same amount of information in both the negative- and positive-toned ads. He found that the tone was not causing the effect at all—it was the amount of information;

negative advertising typically had more information in it than positive advertising, and that was the source of the greater levels of campaign knowledge, interest, and turnout that the other studies had documented.

In reality, cause is rarely **univariate**, or caused by one variable. Social science researchers seldom expect one thing alone to be the direct cause of another. There are usually many things that are responsible. In experiments, this is managed by holding constant other things than the one purposely being varied to make sure those things are not responsible for the effect.

Confounds

Things that could provide plausible alternative explanations are called **confounds**. It is important that there are no confounds—that is, anything that would harm the accuracy of the experiment. Key to a successful experiment is controlling for extraneous influences that might have caused the outcome. For example, a researcher studying the effect of photographs that were placed above or below the fold in a newspaper used as stimuli a vertical photo that showed a person up close above the fold and a horizontal photo showing people far away below the fold. He used *only* the horizontal far-away photo below the fold and *only* the vertical close-up photo above the fold. The problem with this is that it could have been the vertical or horizontal format of the photo, or the close-up or far-away distance of the people in them, that was responsible for the effect he found. Those were not the variables he had deliberately varied. The pictures varied on many levels, not just the ones he was interested in, which was placement in the newspaper. This represents a confound—a plausible alternative explanation that was not controlled for.

To avoid confounds in one study designed to determine how the race of the people in photographs affected journalists' ethical judgment, the researcher took the exact same picture and digitally altered the skin tone, hair, and facial features of the people in order to manipulate their race. That way, everything was the same except race—the thing designed to vary. The backgrounds were the same, the distance from the camera, the people's attractiveness, and everything else was the same. This avoided any confounds. Researchers know, for example, that close-ups make people feel more comfortable with the people in the photographs than long-shots,²⁵ and that attractive people are evaluated better on a variety of characteristics than less attractive people, including trustworthiness and electability in political candidates.²⁶ So this study needed photographs that were exactly the same in every way except the race of the people in them.

In experiments, it is helpful to be as Type A as possible. For example, a researcher might measure the distance from the TV to the chair that subjects sit in before every new

subject comes in. All subjects should be exactly the same distance from the TV because being even a little bit closer can make a difference.

Researchers discover potential confounds two ways: by using common sense and by reading other studies. The literature abounds with evidence showing, for example, that political party and ideology affect many things.²⁷ Thus, many experiments control for these potential confounds by creating stimuli that do not implicitly or explicitly state the political party of a fictitious candidate. This represents an **experimental control**, or carefully managing the variables in play.²⁸ Another example of avoiding potential confounds is not using real issues in the news at the time of an experiment held during an election.²⁹ When a researcher cannot control for a potentially spurious relationship between the independent and dependent variables, another approach is to measure it and **statistically control** for it—that is, take it out of the equation before examining what effect the manipulation had. Measuring people’s political party and ideology, and then using them as covariates, is a form of statistical control.ⁱ

It is important to read the literature to discover things that need controlling. For example, in studies of moral development, researchers have found a connection between being liberal or conservative and quality of moral judgment.³⁰ This is not something that is new, we have known about this for a long time. However, this is key. Age and education are also related to better moral judgment,³¹ so experiments that use moral judgment as the outcome or **dependent variable** typically measure participants’ political ideology in order to incorporate them in the statistical analyses as covariates should random assignment not have already made the groups equivalent on that variable. Covariates work by taking out the effect of the potentially confounding variable so researchers can see the true effect of whatever variable is being manipulated. It is important to know the literature in the domain being studied, because not everything one might suspect as having a potentially spurious relationship between a dependent and independent variable really does. For example, gender does not usually matter in moral judgment,³² but it does in a great many other things. As most effects are rarely caused by simply one thing, it is important to know what else might be affecting the outcome.

It is also important to control for potential confounds that your own common sense tells you might affect the outcome. Just because you do not see something in the literature, if you think it might cause an effect, build it into the experiment so you can

ⁱCovariates are not always necessary in experiments the way they are in observational studies because random assignment is designed to eliminate the effects of such variables. Random assignment will be discussed in more detail in chapter 7.

test it and see. Presumably you are studying a phenomenon that you know a lot about. If you are an expert, you should have some idea of what kinds of things affect the phenomenon you are studying. This is one way that knowledge is created and the literature develops.

Control Groups

Another important requirement of experiments is not just knowing what happens to people who receive the treatment, but also knowing what would happen if they had *not* received the treatment. Having a group of people who do not get the treatment, known as the *control group*, gives us a way to infer what the outcome would be had there been no treatment. The effect is the difference between what *did* happen when people got the treatment and what *would have* happened had they not gotten it. This allows us to isolate and test the effects of one variable at a time, and provides greater certainty that the effect is due to that variable and not something else. The people who are exposed to the treatment or intervention are called the *treatment group*, *experimental group*, or *manipulation group*. The people who are not exposed to the treatment are the control group. This group of people who are not given the manipulation is used as a comparison for what “normal,” “neutral,” or “no manipulation” would be like.

For Review-No Commercial Use(2023)

In medical research, the control group is often given a placebo, which looks like a pill, injection, therapy, or some other treatment but really is only a sugar pill, saline injection, or something else that leads people to perceive they had something done to them. (See More About box 1.3 on placebos.) Because of the power of suggestion, it is not good enough to actually do nothing to the people in the control group; they must perceive that they had some sort of treatment. For social scientists, that can sometimes present a problem. For example, when testing different types of marketing messages, what constitutes a control group? No message, of course, but then what do subjects do instead? You cannot have them just come in and answer the questionnaire without having them be exposed to something or they will perceive the study as too artificial. Control groups will be discussed more in the coming chapters; the point of the discussion here is that experiments need some way to compare what happens to people who get the treatment versus what happens to people who do not.

Assignment

Last in this discussion, but probably most important to experiments, is the topic of how people are assigned to the different interventions in an experiment. This topic is so important that it rates its own chapter (chapter 7). Briefly, the gold standard is

MORE ABOUT . . . BOX 1.3

Placebos

In medical research, the use of placebos is much more complicated than a simple injection of saline, a sugar pill, or the pretense of surgery to mimic an actual treatment on a group of control subjects. In fact, these inert or ineffective but harmless treatments actually have been shown to have effects. Called the *placebo effect*, people receiving them actually report improvement in whatever condition they were supposedly being treated for.

The word *placebo* is Latin for “I shall please.”³³ The first controlled trial to use placebos has been traced to 1801.³⁴ The perception of placebos as fraudulent, deceptive, and unethical arose because up until the middle of the twentieth century, many practicing physicians would administer them to patients under the guise of actual medicine.³⁵ Questions about the ethicality of using placebos in research continue to this day.³⁶



© Jack Conbitt, www.Cartoonstock.com

The first report on the placebo effect found that four out of five patients receiving the placebo reported relief of their symptoms; results were the same when they received the actual treatment.³⁷ The report cites the power of hope, faith, and the imagination. Ever since, research on the placebo effect has continued to show that patients improve after receiving the inert placebo to varying degrees, including equal or better results than the active drug.³⁸ A review of fifteen studies showed that, on average, placebos performed as well as the active treatment 35% of the time.³⁹ Flaws in this review have been pointed out, but research continues to show that placebos “work” about a third of the time.⁴⁰

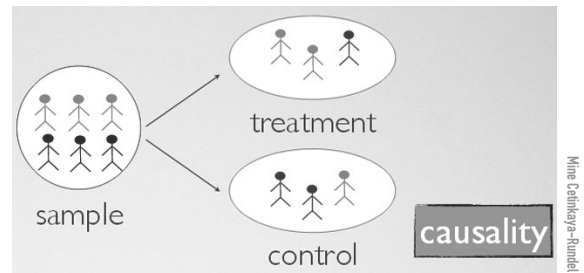
In 1961, Beecher conducted another review and found a placebo effect in 37% of patients.⁴¹ The term placebo effect has been defined as the positive effects of an inert substance due to the power of suggestion,⁴² or “the difference in outcome between a placebo treated group and an untreated control group in an unbiased experiment.”⁴³

In social science research, a placebo is akin to a control group that receives something designed to look like a treatment or manipulation⁴⁴—for example, subjects who read stories about a new movie rather than the episodic and thematically framed crime stories that are the real focus of the study.

A placebo is the inert or fake treatment given to subjects; the placebo effect is subjects’ response to it. It is different from the Hawthorne effect, described in the More About box in chapter 2, which describes how subjects’ performance changes because they are being observed. In the placebo effect, subjects experience change because of their expectations, beliefs, or hopes that the treatment will work rather than the treatment itself.⁴⁵ In order to be effective, a treatment has to pass the placebo test—that is, the treatment must be significantly better on the outcome than the control group that gets the placebo.

Research in the social sciences also finds a placebo effect—for example, with products that promise better athletic performance,⁴⁶ factors that affect financial decisions,⁴⁷ drink labels and perceptions of intoxication,⁴⁸ and well-known brands.⁴⁹ Placebos are also found in daily life—for example, the buttons at intersections that lead pedestrians to feel a sense of control but don’t really affect the light.⁵⁰

random assignment, which is a way of placing subjectsⁱⁱ into the groups in such a way that individual differences are evenly distributed across the different groups. That feature is important to ensure that people's individual characteristics are not confounds of the study. Men and women should be evenly distributed across the groups, as should the young and old, for example. This book devotes an entire chapter to assigning subjects to conditions of experiments, including discussing what to do when it is not possible to assign people this way.



Mine Çelikkaya-Rumel

STARTING A STUDY OF YOUR OWN

Now that we have some basic understanding of what an experiment can do compared to other methods, this chapter turns to the first practical step in starting your own experiment: writing a clear and concise statement of the research problem. Clear and concise writing is important in all kinds of research but especially, in my experience, when writing experiments. Because research can be complex and many social scientists may not be as familiar with experimental methodology as they are with surveys or other techniques, experiments seem to be harder for readers to follow. Thus, one key focus of this book is to help readers write experiments in plain language and terms that anyone with moderate knowledge of social science can understand. That can be harder than you think. We begin with the statement of the research problem. This is more overarching than specific **hypotheses**, although they are related. Writing hypotheses will be developed in a later chapter (chapter 3). To begin a new experiment, one must come up with an idea to test. It helps to think about it in terms of cause and effect, testing if one thing causes another. For example, here are three (rather oversimplified) ideas for studies:

ⁱⁱThis textbook uses the term *subjects* to refer to the people studied in an experiment. The latest (6th) edition of the *Publication Manual of the American Psychological Association* (APA) says both terms, *subjects* and *participants*, are appropriate, noting that “subjects” has been in use for hundreds of years. The history of objections to “subjects” began in 1994 with the fourth edition of the APA manual, when “participants” was preferred. When the current edition appeared in 2010, subjects and participants were on equal footing. The entry on page 73 says, “Indeed, for more than 100 years the term subjects has been used within experimental psychology as a general starting point for describing a sample, and its use is appropriate.” Because subjects is as appropriate as participants, this text uses subjects in order to maintain consistency with other terms in experimental design language, including *between-subjects designs*, *within-subjects designs*, and *human subjects* used by IRBs. For more on this topic, read the essay “What Should They Be Called” by Roddy Roediger in the APS Observer, April 2004, 17, no. 4, available at <http://journal.sjdm.org/roediger.html>.

1. Does seeing a photograph improve ethical reasoning?
2. Does voice pitch affect the credibility of a radio announcer?
3. Does the height of female politicians affect voters' assessments of their qualifications and the likelihood of voting for them?

In the first study, the treatment is showing subjects a photograph versus no photograph in the control condition. In the second study, the voice of a radio announcer was pitched high in one treatment condition and low in another versus normal in the control condition. In the third, pictures of short and tall women were the two treatment groups versus average-height women as the control. Of course, all three experiments ended up being more complex than this, but this simple “does A cause B” approach was the genesis of all the studies.



Writing a Statement of the Problem

The next step is to write a clear and focused statement of the problem to be studied. A good way to start this sentence is with “The purpose of this study is . . .” or something similar. Here are some examples of good, clear, and focused statements of the problem from experiments:

- “This article investigates how media use of the microblogging tool Twitter affects perceptions of the issue covered and the credibility of the information.”⁵¹
- “We experimentally study the common wisdom that money buys political influence.”⁵²
- “We assess the extent to which communication setting (i.e., face-to-face versus online chat room discussion) affects individuals’ willingness to express opinions.”⁵³
- “A first question is whether providing general information on the welfare properties of prices and markets modify attitudes toward repugnant trades.”⁵⁴

- “. . . the goal of this study is to demonstrate that moral convictions and moral judgments in politics are causally affected by harm associations and moral emotions.”⁵⁵

All of these examples have three things in common: They say what the intervention, manipulation, or cause is, and what the effect or outcome is, usually in that order. And they have a verb. In the first example, Twitter is the intervention or cause, and the effect or outcomes are perceptions of the issue and credibility. In the next one, money is the cause, and political influence is the effect. In the last example, the order is reversed, with the effect—moral convictions and judgments—listed first, and the causes—harm associations and moral emotions—given last. You may also recognize these as independent variables (the cause, interventions, or manipulations) and dependent variables (the effect or outcomes). The third thing in common is that there is a verb in each of these, some word that describes the action the cause is expected to have on the outcome, or what one thing is expected to do to the other. In these examples, the verbs are “affects,” “buys,” “modify,” and “causally affected.” For experiments, it is advisable to stay away from less precise words such as “explore,” “understand,” and “examine,” and instead to use more specific, causal language such as this. Other good words to use include “differs,” “improves,” and “influences.” For example, a statement might say, “The purpose of this study is to test the idea that photographs *improve* ethical reasoning,” or “The purpose of this study is to see if assessments of credibility *differ* with the pitch of a radio announcer’s voice.”

To write a clear and focused statement of the research problem for an experiment, first determine what the cause (or intervention, manipulation, independent variable) is, then say what it is expected to do (differ, affect, modify, cause—the verb) to some outcome (the effect or dependent variable). Here, I offer my fill-in-the-blanks template for writing a statement of the research problem:

“*The purpose of this study is to see how* (one thing, insert the cause, intervention, manipulation, treatment, or independent variable) (does something, insert a verb—differs, affects, modifies, causes, changes, etc.) *to* (something else, insert effect, outcome, dependent variable).”

Some of the terms in these examples, such as *harm associations*, might not be familiar if they are outside your discipline, so it helps to examine some experiments in your own field for other examples that may be more commonplace. Also notice how all of these are

only one sentence long. Writing such a clear, focused, and easily understandable statement that describes the experiment in one sentence is not easy. It is not surprising to have to write, rewrite, and edit it many times. Have someone familiar with the discipline read it and see if he or she can understand it. As the project evolves, you may have to rewrite this statement, maybe even a few times.

Many experiments have more than one purpose, so two or three statements of the problem may be written. If that is the case, put them together in the paper and link them with phrases like “This study also seeks to . . .” For example: “This study tests whether photographs improve ethical reasoning. If this effect is found, this study also seeks to determine the causal mechanism for this improvement.” This way, one has the primary purpose of the experiment followed by additional purposes all in one place rather than spread out around the paper. Readers will appreciate having everything the study intends to do all listed together rather than reading a slowly evolving purpose of the study, waiting for it to unfold like a murder mystery.

Once satisfied with the statement of the problem, authors will need to remind readers periodically throughout the paper what the mission is, so be sure that the statement of the problem stays consistent. One frequent problem I see when reviewing experiments for journals is that the statement of the study changes as the paper progresses. Finally, a good practice is to write the statement of the problem on a sticky note and paste it on your computer where you can see it as you work. This will help keep you focused and consistent.

Answering the “So What” Question

After formulating this clear and focused statement, the next step is to articulate why this study is important. This is commonly referred to as answering the “so what” question. Some journals even have a highlighted box that is devoted to this—for example, see the “Significance” box in *Proceedings of the National Academy of Sciences*.⁵⁶ This is another area where researchers frequently think the importance of their study is obvious and should not have to point it out; in fact, this is one of the most crucial aspects of a study. For this task, I tell students they need to state the obvious. It may be abundantly clear to you, but it will not necessarily seem so to other readers. To do this, think about the reason your study is important on three different levels: (1) to other academics, (2) to professionals in the field, and (3) to society or people in general. For the first, you can point out some gap in knowledge or some obstacle that the study overcomes. It should

also contribute to theory in some way or help uncover any of the mechanisms or reasons for some phenomenon. One reason that is never acceptable alone is because a study has never been done before. That is a start, but you should always go further to say why it is important that it be done beyond never having been done before; otherwise, perhaps it has never been done before for a good reason.

Frequently, one sees researchers point out the importance to other scholars but overlook the other two groups: professionals and the public. For any kind of study, being able to articulate why your findings are meaningful beyond academia is crucial to research that makes a difference. R. Barker Bausell's book *Conducting Meaningful Experiments*⁵⁷ is predicated on that premise. For the second step in considering the “so what” question, ask if your study will be of interest to those in the professional arm of your discipline. Will it help accountants, campaign managers, teachers, public relations professionals, or anyone else do their jobs better? Will the findings of the study give professionals more insight into their own subconscious decision making? Will it help them overcome some obstacle or give them evidence they need to change the way they practice their craft? Perhaps it will show them which of their efforts are paying off in the outcome they desire and which are not, whether that is more engaged citizens or more customers. Making this kind of concerted effort to solve real-world problems helps bridge the gap between scholars and the profession they serve.

For Review-No Commercial Use(2023)

Finally, being able to say why all of society will benefit results in more meaningful science. Too often, the public considers academic researchers to be “eggheads in ivory towers” writing about things that have no basis in reality in order to get another publication. There are even awards that make the news for the most wasteful research.⁵⁸ Having one's research ridiculed in this way might be avoided if studies better articulated why seemingly silly or obvious findings are important to someone other than our scholarly colleagues. This kind of attention does nothing to help advance the cause of research or increase funding for it—conducting research that is meaningful to ordinary people does. So answer the “so what” question; will it help anyone or improve any social ills? Not every study will broker world peace, but it might help lessen racial profiling, change a morally repugnant practice, or give policy makers information needed to pass a law. Not every study articulates all three, but thinking through the benefits to these different publics can help experimentalists design studies that are truly meaningful. For some examples of statements that answer the “so what” question, see How To Do It box 1.4.

HOW TO DO IT 1.4

Examples of Answers to the “So What” Question

From: Neil, Nicole, and Emily A. Jones. 2015. “Studying Treatment Intensity: Lessons from Two Preliminary Studies.” *Journal of Behavioral Education* 24: 51–73.

“There is only a recent and small literature examining treatment intensity, and the research on treatment intensity focused on specific disorders is even more limited. It may be that etiology and characteristics associated with specific etiologies impact the effects of intervention intensity. Many children with Down syndrome display poor task persistence and inconsistent motivational orientation . . . For some learners with Down syndrome, it is possible that there is an optimum moderate level of intensity, past which learners engage in greater levels of escape-motivated problem behavior and there are diminishing gains in acquisition rates.”⁵⁹

From: Coleman, Renita. 2011. “Color Blind: Race and the Ethical Reasoning of African Americans on Journalism Dilemmas.” *Journalism and Mass Communication Quarterly* 88 (2) (Summer): 337–351.

“This study is of value because it provides important information to evaluate one of the solutions offered to the problem of stereotypical media portrayals—hiring and promoting more minority journalists. Newsrooms across the country are staffed primarily by whites; incorporating more minority viewpoints should lead to more equal coverage of minorities, according to the argument. . . . it is important to examine whether minority journalists do in fact exhibit more tolerant attitudes toward minorities in their cognitive processing. To date, there is no empirical evidence that black journalists have more favorable perceptions of blacks in the news. . . . This study also fills that void by exploring how race influences the ethical reasoning of blacks when blacks and whites are in news stories.”⁶⁰

From: Aday, Sean. 2006. “The Framesetting Effects of News: An Experimental Test of Advocacy Versus Objectivist Frames.” *Journalism and Mass Communication Quarterly* 83 (4) (Winter): 767–784.

“Specifically, we still do not know enough about why effects are found in some cases and not others, and too little work has been done exploring the cognitive basis of the effect that would allow us to develop a theory for when and why some attributes would have a second-level effect and others would not.”⁶¹

From: Elias, Julio J., Nicola Lacetera, and Mario Macis. 2015. “Markets and Morals: An Experimental Survey Study.” *PLoS ONE* 10 (6) (June 1): 1–13. Public Library of Science.

“Prohibiting some of these transactions has costs. Life insurance contracts, for instance, were once illegal because they were seen as gambles against God; they now create value for millions of people, and are viewed as a form of ‘institutionalized altruism.’ Similarly, the idea of an all-volunteer paid army was long rejected in the United States, despite arguments showing its efficiency. The prohibition of payments to people who give their organs contributes to the growing gap between organ demand and supply. Banning some trades may also lead to the formation of illegal markets, which, in turn, entail further costs such as violence . . .”⁶²

From: Grober, Jens, Ernesto Reuben, and Agnieszka Tymula. 2013. "Political Quid Pro Quo Agreements: An Experimental Study." *American Journal of Political Science* 57 (3) (July): 582–597.

"There are good reasons to suspect there is some truth behind the common belief that money in politics is undesirable. First, in spite of being banned, political quid pro quo can occur outside publicly observable channels. Second, for economically powerful special interests, most of which are large corporate firms, giving as an investment that increases profits is a more plausible explanation than political participation. Moreover, returned favors to such interests, such as specific tax breaks, subsidies, and regulations, can be easily concealed as an economic necessity and are therefore hard to quantify. Third, collusion between major candidates may also take the form of an agreement on a common view with regard to a given political issue . . . Finally, even if the impact of money in politics is overestimated by the public, this belief can affect the public's political trust and behavior."⁶³

Common Mistakes

- Not clearly stating the purpose of the research, and not keeping it consistent throughout the paper
- Not putting all the things the study intends to do in one place, but having the goals of the study unfold slowly through the article
- Not putting the statement of purpose up high in the paper, in the introduction and before page 3
- Failing to say why the study is important to theory, other researchers, the profession, and regular people. "Because it has never been done" is not a reason why a study is important by itself.

Test Your Knowledge

1. Experiments need to show that the cause precedes the effect but not that it is necessarily related to it. If there is a statistically significant relationship between two things, that is all that matters.
 - a. True
 - b. False
2. A researcher studied the effects of attractiveness on how well students liked a teacher. The attractive teacher was twenty-five years old; the unattractive teacher was forty-five years old. The problem here is that:

- a. Age is confounded with attractiveness
 - b. It is hard to define attractiveness
 - c. The cause did not precede the effect
 - d. The researcher did not control for political ideology
3. You assign fifteen employees to go to a one-day seminar on stress management. Another fifteen are assigned to a one-week seminar. At the end of the month, you measure each employee's perceived level of stress. What is the treatment or manipulation in this study?
- a. How stressed out employees are
 - b. How long the seminar is
 - c. How you chose the thirty employees
 - d. The quality of the stress management teacher
4. A study measures students' arousal level before they take a test. It finds that as arousal increases, performance decreases. This finding shows:
- a. Causality
 - b. Correlation
 - c. A plausible alternative explanation
 - d. A confound
5. Which of the following is NOT one of the three basic criteria for an experiment?
- a. Cause must precede the effect
 - b. The effect must be unlikely to have occurred by chance
 - c. Cause must be related to the effect
 - d. There are no plausible alternative explanations for the effect
6. Variation is achieved by:
- a. Holding everything constant
 - b. Using demographics as covariates
 - c. Systematically changing something
 - d. Having a control group
7. Things that could provide plausible alternative explanations are called:
- a. Covariates
 - b. Confounds
 - c. Independent variables
 - d. Causal mechanisms

For Review-No Commercial Use(2023)

8. In experiments, control groups serve the purpose of:
 - a. Allowing us to know what happens to people who receive the treatment
 - b. Allowing us to generalize to more people
 - c. Allowing us to know what would happen to subjects if they had not received the treatment
 - d. Allowing us to say some effect occurred with a specific degree of certainty
9. In an experiment, “assignment” is:
 - a. The task that subjects must complete
 - b. How authorship order is calculated for the paper
 - c. The way researchers ensure subjects believe the experiment is real
 - d. How subjects are put in the different interventions or control group
10. The gold standard in experiments is to assign subjects:
 - a. Representatively
 - b. Purposively
 - c. Randomly
 - d. Haphazardly

For Review-No Commercial Use(2023)

Answers:

- | | | | |
|------|------|------|-------|
| 1. b | 4. b | 7. b | 9. d |
| 2. a | 5. b | 8. c | 10. c |
| 3. b | 6. c | | |

Application Exercises

1. Use scholar.google.com or your school library’s database to find studies that use experimental designs in your discipline. Include the word *experiment*, *experimental design*, or *controlled experiment* in the search terms. Read three of the experiments that interest you most and look for the concepts covered here. Specifically, identify the treatment group or groups. Is there a control group? If so, what is used to represent “no treatment”? Identify the statement of the problem and the answer to the “so what” question.
2. Think up three distinct studies that you would like to do with an experiment. That is, something should be manipulated or changed in order to see what effect it has on some outcome. Write a clear and focused statement of the problem. Explain why it is important to academics, the profession, and the world at large (the “so what” question). Use 250 words for each. These should not be straight replications but new ideas, or they may be replications with substantial extensions to the study you are replicating.

Suggested Readings

- The introduction and Chapters 1 and 2 of Bausell, R. B. 1994. *Conducting Meaningful Experiments: 40 Steps to Becoming a Scientist*. Thousand Oaks, CA: Sage.
- Chapter 1 in Shadish, W. R., T. D. Cook, and D. T. Campbell. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Belmont, CA: Wadsworth Cengage Learning.
- Thorson, Esther, Robert H. Wicks, and Glenn Leshner. 2012. "Experimental Methodology in Journalism and Mass Communication Research." *Journalism and Mass Communication Quarterly* 89 (1): 112–124.

Notes

1. Markus MacGill, "What Is a Randomized Controlled Trial in Medical Research?" *Medical News Today*, <http://www.medicalnewstoday.com/articles/280574.php>. Accessed February 10, 2017.
2. Harald O. Stolberg, Geoffrey Norman, and Isabelle Trop, "Fundamentals of Clinical Research for Radiologists: Randomized Controlled Trials," *American Journal of Roentgenology* 183, no. 6 (2004): 1339–1344.
3. Marcia L. Meldrum, "A Brief History of the Randomized Controlled Trial: From Oranges and Lemons to the Gold Standard," *Hematology/Oncology Clinics of North America* 14, no. 4 (2000): 745–760.
4. S. Shikata et al., "Comparison of Effects in Randomized Controlled Trials with Observational Studies in Digestive Surgery," *Annals of Surgery* 244, no. 5 (2006): 668–676; Stolberg, Norman, and Trop, "Fundamentals of Clinical Research for Radiologists."
5. Streptomycin in Tuberculosis Trials Committee, "Streptomycin Treatment of Pulmonary Tuberculosis: A Medical Research Council Investigation," *British Medical Journal* 2, no. 4582 (1948): 769–782.
6. D. T. Campbell and J. C. Stanley, *Experimental and Quasi-Experimental Designs for Research* (Chicago: Rand McNally, 1963), 3.
7. Esther Thorson, Robert H. Wicks, and Glenn Leshner, "Experimental Methodology in Journalism and Mass Communication Research," *Journalism and Mass Communication Quarterly* 89, no. 1 (2012): 112–124.
8. R. McDermott, "New Directions for Experimental Work in International Relations," *International Studies Quarterly* 55, no. 2 (2011): 503–520.
9. Matthew C. Makel et al., "Replication of Special Education Research: Necessary but Far Too Rare," *Remedial and Special Education* 37, no. 3 (2016): 1–8.
10. Rebecca B. Morton and Kenneth C. Williams, *Experimental Political Science and the Study of Causality: From Nature to the Lab* (New York: Cambridge University Press, 2010): 5–6, 16–17; Gerry Stoker, "Exploring the Promise of Experimentation in Political Science: Micro-Foundational Insights and Policy Relevance," *Political Studies* 58 (2010): 300–319; Cengiz Erisen, Elif Erisen, and Binnur Ozkececi-Taner, "Research Methods in Political Psychology," *Turkish Studies* 13, no. 1 (2013): 13–33; James N. Druckman et al., "The Growth and Development of Experimental Research in Political Science," *American Political Science Review* 100, no. 4 (2006): 627–635.
11. Ulrich Hamenstädt, "Teaching Experimental Political Science: Experiences from a Seminar on Methods," *European Political Science* 11, no. 1 (2012): 114–127; Erisen, Erisen, and Ozkececi-Taner, "Research Methods in Political Psychology"; James N. Druckman and Arthur Lipia, "Experimenting with Politics," *Science* 335, no. 6 (March 9, 2012): 1177–1179; Luke Keele, Corrine McCaughy, and Ismail White, "Strengthening the Experimenter's Toolbox: Statistical Estimation of Internal Validity," *American Journal of Political Science* 56, no. 2 (2012): 484–499; Sharon Crasnow, "Natural Experiments and Pluralism in Political Science," *Philosophy of the Social Sciences* 45, no. 4/5 (2015): 424–441.
12. Yair Levy, Timothy J. Ellis, and Eli Cohen, "A Guide for Novice Researchers on Experimental and

- Quasi-Experimental Studies in Information Systems Research,” *Interdisciplinary Journal of Information, Knowledge and Management* 6 (2011): 151–161.
13. Susan D. Hyde, “Experiments in International Relations: Lab, Survey, and Field,” *Annual Reviews of Political Science* 18 (2015): 403–424. Hamenstädt, “Teaching Experimental Political Science.”
 14. Erisen, Erisen, and Ozkececi-Taner, “Research Methods in Political Psychology.”
 15. William J. Reid, Bonnie Davis Kenaley, and Julianne Colvin, “Do Some Interventions Work Better Than Others? A Review of Comparative Social Work Experiments,” *Social Work Research* 28, no. 2 (2004): 71–81.
 16. Joshua B. Plavnick and Summer J. Ferreri, “Single-Case Experimental Designs in Educational Research: A Methodology for Causal Analyses in Teaching and Learning,” *Education Psychology Review* 25 (2013): 549–569.
 17. William R. Shadish, Thomas D. Cook, and Donald T. Campbell, *Experimental and Quasi-Experimental Designs for Generalized Causal Inference* (Belmont, CA: Wadsworth Cengage Learning, 2002).
 18. K. Imai, G. King, and E. Stuart, “Misunderstandings Between Experimentalists and Observationalists About Causal Inference,” *Journal of the Royal Statistical Society Series A* 171 (2008): 481–502.
 19. Gustav Fischer, “Ornithologische Monatsberichte,” *Jahrgang*, 44, no. 2 (1936); Gustav Fischer, “Statistisches Jahrbuch Deutscher Gemeinden,” *Jahrgang*, 48, no. 1 (1940): 27–33, <http://pignottia.faculty.mjc.edu/math134/classnotes/storks.pdf>
 20. <http://faculty.vassar.edu/lowry/ch3pt2.html>
 21. Carter Hay et al., “Inside the Black Box: Identifying the Variables That Mediate the Effects of an Experimental Intervention for Adolescents,” *Crime & Delinquency* 61, no. 2 (2015): 243–270.
 22. Renita Coleman, “The Effect of Visuals on Ethical Reasoning: What’s a Photograph Worth to Journalists Making Moral Decisions?” *Journalism and Mass Communication Quarterly* 83, no. 4 (2006): 835–850.
 23. H. Denis Wu and Renita Coleman, “The Affective Effect on Political Judgment: Comparing the Influences of Candidate Attributes and Issue Congruence,” *Journalism and Mass Communication Quarterly* 91, no. 3 (2014): 530–543.
 24. Daniel Stevens, “Tone Versus Information: Explaining the Impact of Negative Political Advertising,” *Journal of Political Marketing* 11, no. 4 (2012): 322–352.
 25. Arthur Asa Berger, “Semiotics and TV,” in *Understanding Television: Essays on Television as a Social and Cultural Force*, ed. R. R. Adler (New York: Praeger, 1981): 91–114.
 26. Zhao Na et al., “Face Attractiveness in Building Trust: Evidence From Measurement of Implicit and Explicit Responses,” *Social Behavior and Personality: An International Journal* 43, no. 5 (2015): 855–866; Leslie A. Zebrowitz, Robert G. Franklin Jr., and Rocco Palumbo, “Ailing Voters Advance Attractive Congressional Candidates,” *Evolutionary Psychology* 13, no. 1 (2015): 16–28.
 27. Amy E. Lerman, Katherine T. McCabe, and Meredith L. Sadin, “Political Ideology, Skin Tone, and the Psychology of Candidate Evaluations,” *Public Opinion Quarterly* 79, no. 1 (2015): 53–90; Sarah Reckhow, Matt Grossmann, and Benjamin C. Evans, “Policy Cues and Ideology in Attitudes toward Charter Schools,” *Policy Studies Journal* 43, no. 2 (2015): 207–227; Jonathon Schultdt and Adam Pearson, “The Role of Race and Ethnicity in Climate Change Polarization: Evidence from a U.S. National Survey Experiment,” *Climatic Change* 136, no. 3/4 (2016): 495–505.
 28. Stoker, “Exploring the Promise of Experimentation in Political Science.”
 29. Kevin J. Mullinix et al., “The Generalizability of Survey Experiments,” *Journal of Experimental Political Science* 2 (2015): 109–138.
 30. J. R. Rest and Darcia Narvaez, eds., *Moral Development in the Professions: Psychology and Applied Ethics* (Hillsdale, NJ: Erlbaum, 1994).
 31. James. R. Rest, “Morality,” in *Handbook of Child Psychology, Vol. III Cognitive Development*, ed. P. H. Mussen (New York: Wiley, 1983).
 32. Stephen Thoma, “Estimating Gender Differences in the Comprehension and Preference of Moral Issues,” *Developmental Review* 6, no. 2 (1986).
 33. A. J. De Craen et al., “Placebos and Placebo Effects in Medicine: Historical Overview,” *Journal of the Royal Society of Medicine* 92, no. 10 (1999): 511–515.
 34. Ibid.
 35. Ibid.; David B. Elliott, “The Placebo Effect: Is It Unethical to Use It or Unethical Not To?” *Ophthalmic & Physiological Optics* 36, no. 5 (2016): 513–518.

36. Elliott, "The Placebo Effect."
37. H. K. Beecher, "The Powerful Placebo," *JAMA* 159, no. 17 (1955): 1602–1606.
38. De Craen et al., "Placebos and Placebo Effects in Medicine."
39. Beecher, "The Powerful Placebo."
40. De Craen et al., "Placebos and Placebo Effects in Medicine."
41. H. K. Beecher, "Surgery as Placebo: A Quantitative Study of Bias," *JAMA* 176 (1961): 1102–1107.
42. Beecher, "The Powerful Placebo."
43. P. C. Goetzsche, "Is There Logic in the Placebo?" *Lancet* 344, no. 8927 (1994): 904.
44. Alan Bryman, "Placebo," in *The Sage Encyclopedia of Social Science Research Methods*, ed. Michael S. Lewis-Beck, Alan Bryman, and Tim Futing Liao (Thousand Oaks, CA: Sage, 2011): 825.
45. Ibid.
46. Aaron M. Garvey, Frank Germann, and Lisa E. Bolton, "Performance Brand Placebos: How Brands Improve Performance and Consumers Take the Credit," *Journal of Consumer Research* 42, no. 6 (2016): 931–951.
47. Michael A. Kuhn, Peter Kuhn, and Marie Claire Villeval, "Decision-Environment Effects on Intertemporal Financial Choices: How Relevant Are Resource-Depletion Models?" *Journal of Economic Behavior & Organization* 137 (2017): 72–89.
48. Yann Cornil, Pierre Chandon, and Aradhna Krishna, "Does Red Bull Give Wings to Vodka? Placebo Effects of Marketing Labels on Perceived Intoxication and Risky Attitudes and Behaviors," *Journal of Consumer Psychology (Elsevier Science)* 27, no. 4 (2017): 456–465.
49. Carlos Alberto Alves, Evandro Luiz Lopes, and José Mauro da Costa Hernandez, "It Makes Me Feel So Good: An Experimental Study of the Placebo Effect Generated by Brands," *Journal of International Consumer Marketing* 29, no. 4 (2017): 223–238.
50. Gwen Sharp, "The Placebo Effect," *The Society Pages*, March 10, 2011, <https://thesocietypages.org/socimages/2011/03/10/the-placebo-effect/>.
51. Mike Schmierbach and Anne Oeldorf-Hirsch, "A Little Bird Told Me, So I Didn't Believe It: Twitter, Credibility, and Issue Perceptions," *Communication Quarterly* 60, no. 3 (2012): 317–337.
52. Jens Grober, Ernesto Reuben, and Agnieszka Tymula, "Political Quid Pro Quo Agreements: An Experimental Study," *American Journal of Political Science* 57, no. 3 (2013): 582–597.
53. Shirley S. Ho and Douglas M. McLeod, "Social-Psychological Influences on Opinion Expression in Face-to-Face and Computer-Mediated Communication," *Communication Research* 35, no. 2 (2008): 190–207.
54. Julio J. Elias, Nicola Lacetera, and Mario Macis, "Markets and Morals: An Experimental Survey Study," *PLoS ONE* 10, no. 6 (2015): 1–13.
55. Pazit Ben-Nun Bloom, "Disgust, Harm and Morality in Politics," *Political Psychology* 35, no. 4 (2014): 495–513.
56. For an example, see William Minozzi et al., "Field Experiment Evidence of Substantive, Attributional, and Behavioral Persuasion by Members of Congress in Online Town Halls," *Proceedings of the National Academy of Sciences* 112, no. 13 (2015): 3937–3942.
57. R. Barker Bausell, *Conducting Meaningful Experiments: 40 Steps to Becoming a Scientist* (Thousand Oaks, CA: Sage, 1994).
58. FoxNews.com, "Senate Report Finds Billions in Waste on Science Foundation Studies," Foxnews.com, <http://www.foxnews.com/politics/2011/05/26/senate-report-finds-billions-waste-science-foundation-studies/>.
59. Nicole Neil and Emily A. Jones, "Studying Treatment Intensity: Lessons From Two Preliminary Studies," *Journal of Behavioral Education* 24 (2015): 51–73.
60. Renita Coleman, "Color Blind: Race and the Ethical Reasoning of African Americans on Journalism Dilemmas," *Journalism and Mass Communication Quarterly*, 88, no. 2 (Summer 2011): 337–351.
61. Sean Aday, "The Framesetting Effects of News: An Experimental Test of Advocacy Versus Objectivist Frames," *Journalism and Mass Communication Quarterly* 83, no. 4 (2006): 767–784.
62. Elias, Lacetera, and Macis, "Markets and Morals," 2.
63. Grober, Reuben, and Tymula, "Political Quid Pro Quo Agreements," 582–583.



ETHICS AND FAMOUS EXPERIMENTS IN HISTORY

*If I have seen further than others, it is by standing upon the shoulders
of giants.*

—Isaac Newton

For Review-No Commercial Use(2023) LEARNING OBJECTIVES

- Locate the historical origins of methods, techniques, and concepts used in experimental research today.
- Illustrate the importance of creativity to solving research problems.
- Summarize the unintended consequences of experimental research as a foundation for the development of ethical guidelines and Institutional Review Boards.
- Explain the harmful effects of using deception in experiments.
- Describe how to use deception ethically in the context of experiments.

This chapter briefly deviates from the practical how-to approach in order to give some context in which to understand the role of current issues, concepts, and techniques in social science experiments. This brief history will illustrate how experimental methodology has developed by using stories of some of the most famous, and infamous, experiments from

the past—for example, how **randomization**, or the process of assigning subjects to groups, came about, and why experimentalists use terms like *split plots*. Some stories highlight ingenious ways to answer questions, encouraging researchers to think creatively about their own studies. Creativity is important in research not just for novelty's sake but also to help answer questions in the best way possible. Other examples show the importance of protecting participants from harm and how they helped lead to the development of Institutional Review Boards (IRBs), committees charged with ensuring researchers follow ethical guidelines. Many of these stories are quite entertaining, and some have even been made into movies for mass audiences. More important, understanding the historical roots can lead to a better grasp of contemporary practices. Other examples show how findings from experiments can affect real-world problems—for example, the famous studies that examined why good people do bad things. In this chapter are some of the stories of the “giants” from the quote by Isaac Newton upon whose shoulders experimental methodology is built.

The history of experimental design is filled with fascinating studies, often hyped as the most evil, creepy, bizarre, or ones that went horribly wrong. Studies that could never be conducted again are also a popular theme in the “mad science” category. This chapter will take a different approach; instead of dwelling on the findings or theories created by these experiments, it will highlight the development and early use of important discoveries such as random assignment, controls, and the use of confederates. Importantly, how ethical guidelines developed and IRBs came about will be included. (Chapter 11 will discuss IRBs and other ethical issues not fully addressed here in more detail.) Many of these studies are already known through popular folklore; in those cases, I try not to repeat the obvious but instead highlight aspects that are not as familiar. This chapter does not purport to be exhaustive or cover every important experiment ever conducted; of necessity, many are left out. In particular, the received history tends to leave out women and experimentalists of color. This chapter deals with some of these forgotten figures in More About . . . box 2.1, Contributions of Women. Many historians trace the development of experiments back to the ancient Greeks or others. After the introduction of one very early and important medical experiment, I devote the rest of this chapter to more modern social science experiments.

THE SCURVY STUDIES

Some of the earliest experiments were conducted on a disease called scurvy in 1747. Rare today, it was particularly problematic on ships as it caused sailors to become weak and anemic, and also caused their skin to bleed and gums to rot.²³ A ship's surgeon, James Lind, carried out one of the first controlled experiments to find a cure.²⁴ He chose twelve men who all had scurvy, using only men who “were as similar as I could have them.”²⁵ He then divided them into six groups, putting two men in each. These six

MORE ABOUT . . . BOX 2.1

Contributions of Women

While men are most frequently cited as early pioneers of experimental design in the social sciences, there were also many important women, including women of color. These scientists faced barriers and discrimination in their careers, including a lack of fellowships and being barred from admission to graduate programs and employment in academic positions that allowed research and publishing.¹ Some women completed doctoral work, including theses and dissertations, but were denied the degrees. Men received credit for some of the contributions of women. Some women collaborated with their husbands, and antinepotism policies prevented them from being hired. Racial discrimination placed even more burdens on minority women.²

What few histories include women primarily note their theoretical contributions rather than their advancements in the methodology of experimental design, which is the subject of this book.³ Next are short profiles of three women, including one African American, who made contributions to experimental methodology.⁴ More certainly deserve recognition.

Mary Whiton Calkins is one of the pioneers of experimental psychology, remembered for, among other things, inventing the paired-associates task, a test of memory using paired numbers and colors.⁵ Despite having completed all the requirements for a doctorate, Harvard refused to grant her a degree in 1890 because she was a woman.⁶ Yale and the University of Michigan offered her admission, but she turned them down because they lacked a laboratory for experiments, which Harvard had. It was easier for women to find academic positions at women's colleges, so Calkins went to work at Wellesley and established an experimental laboratory there. Despite not having a doctorate, Calkins published four books and more than 100 papers in scientific journals.⁷ Her accomplishments led to Columbia University and Smith College awarding her honorary doctorates. Despite this recognition and a petition to Harvard signed by thirteen graduates who were prestigious alumni, she was again denied a degree in 1927.⁸

Calkins crossed paths with another experimentalist in this chapter, Joseph Jastrow, who along with Charles S. Peirce helped identify the benefits of random assignment. Jastrow published a study on the kinds of words produced by men and women when asked to write them out quickly, concluding that women's words were repetitive, individual, and concrete, whereas men's words were constructive, useful, and abstract.⁹ She criticized his conclusion, pointing out the confounding effects of the environment, training, and socialization of women.¹⁰ She was an advocate for women's rights all her life, repeatedly refusing to accept a doctoral degree from Radcliffe for work she did at Harvard.

Mamie Phipps Clark was an African American psychologist noted for developing the Clark Doll Test for her research on race, which was used in the 1954 *Brown v. Board of Education* case that allowed African American students to attend White schools.¹¹ Unlike Calkins, Clark earned a doctoral degree in 1943, becoming the first African American woman to do so from



William Notman



New York Post Archives/Contributor

(Continued)

(Continued)

Columbia University. The first African American male to earn a doctorate from Columbia was her husband and research partner, Kenneth B. Clark.¹² He always credited her with the idea for the doll test.¹³

Mamie Clark's contribution to experimental methods involved showing children two identical dolls, one Black and one White. The children were asked which doll was bad and which was good, which one the child liked to play with, and which one most looked like them. Many Black children identified the Black doll as bad, and almost half said the White doll looked most like them. This was more pronounced in children from segregated schools than integrated ones.¹⁴ The Clarks' experiments were important evidence that segregation harmed children, and also were influential in the first mass-produced doll of a Black infant.¹⁵ Clark never found a position in academia, instead working as a researcher and then clinical psychologist at a children's home until she and her husband opened a testing and consultation center for minority children in Harlem.¹⁶



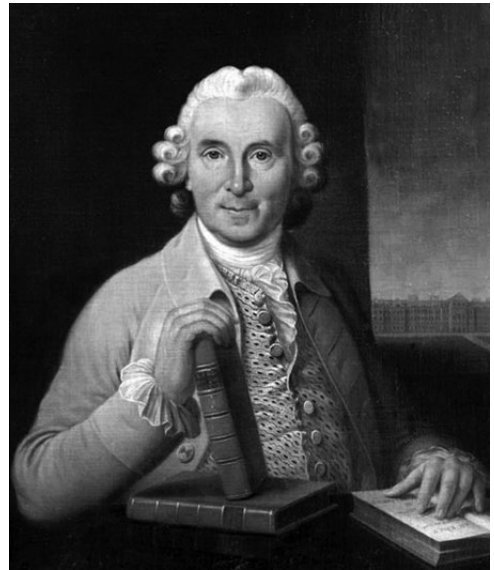
JHU Sheridan Libraries/Gard/Contributor

Mary Ainsworth is remembered for an assessment technique known as the "Strange Situation."¹⁷ Work using this method helped advance psychologists' understanding of children's attachment to their caregivers. Ainsworth's method involved researchers observing through a one-way mirror a child's behavior during eight different episodes of about three minutes each where a mother and child come to the researcher's laboratory, which is filled with toys. The different episodes involve a stranger entering the room and trying to befriend the child, the mother leaving the child alone with the stranger, returning, both mother and stranger leaving the child alone, the stranger re-entering to comfort the child, and finally, the stranger picking the child up.¹⁸ Observations were recorded every 15 seconds on a 1 to 7 scale. The assessment had good reliability, meaning other researchers could reproduce the findings.¹⁹ Typical for a laboratory experiment, the method was criticized as being artificial and lacking in ecological validity,²⁰ and also on ethical grounds for causing stress in young children.²¹

Many other women also contributed to experimental methodology by authoring or coauthoring other tests and measures, including Grace Kent-Rosanoff's Word Association Test, Florence Goodenough's Draw-a-Man Test, and Grace Fernald's character tests that preceded other tests of moral development, among many others.²²

groups then got different treatments ranging from a quart of cider; sulphuric acid; a half pint of seawater; a mixture of garlic, mustard, and horseradish in a dose "the bigness of a nutmeg"²⁶; vinegar; or two oranges and a lemon.²⁷ Lind had six treatment conditions for this medical study, which is considered a lot in social science today, with two to four more common.²⁸ He also had a very small number of subjects—only twelve, with just two in each treatment group. Small sample sizes are more typical in medical studies, where large effects are more common than in social science. Another important issue in this study was the problem of how to assign people to conditions. Lind recognized the dilemma with groups where people were dissimilar on important factors, understanding that people's inherent individual differences, such as age, weight,

and general health, could affect the results. He limited his sample to men who were as similar as possible on certain requirements in order to get results of the treatment that were not confounded by these extraneous factors. Here, he used what would later be known as a **matching** strategy, pairing up men with similar characteristics. Importantly, he manipulated only these six conditions while holding everything else (that he could think of) constant. His independent variables—the six different dietary supplements—were not drawn from theory but from cures that had been proposed earlier. However, Lind did write about how the theory on scurvy was mostly conjecture by researchers who had never seen it, and advocated a mix of theory and hands-on experience. His measurement of one of the treatments' doses—the size of a nutmeg—is not very precise; today, that would be measured in grams or something similar.



James Lind

Wikimedia Commons

Lind's study was a huge success by any standard. After only six days, the men who were given the oranges and lemon recovered.²⁹ The other men also improved, but the ones who ate citrus fruit had a dramatic recovery compared to the others, which eventually led to the recognition that vitamin C was the agent at work. No statistical tests of significance were performed, but the effects at six days were obvious.

This study was important for several reasons; it recognized that people needed to be assigned to treatment groups in such a way that individual differences would not matter, which Lind accomplished by carefully selecting men who had similar characteristics. This was in 1747. By the late 1800s, assigning subjects to conditions had become a burning topic and a different solution emerged, although it was applied to the treatments rather than the subjects, as illustrated in the next story.

THE CONTRIBUTIONS OF CHARLES PEIRCE

Charles Sanders Peirce and his student Joseph Jastrow were by some accounts the first to use random assignment.³⁰ The first study to use it was conducted in 1885 to see if people could judge how much something weighed just by feeling and looking at it.³¹ The theoretical construct Peirce (pronounced "purse") was interested in was the source of judgment errors, resulting in concepts such as the just-noticeable difference, which business, marketing, and advertising scholars will recognize. In the first study, Peirce began the first experiment by always starting and ending with the heaviest weights. In the second study, he tried alternating the heaviest with the lightest weight. Last, he describes how he used a pack of cards



Charles Sanders Peirce

to randomly assign the order of the weights. His results were vastly different from the first two trials when the order of the weights had not been randomized, showing the importance of this technique.³² Peirce used a deck of cards and simply shuffled them before drawing out a card that would determine which weight would come next. Today, we tend to use random number generators, but a deck of cards will still work just as well. He talked about how this method occasionally produced “long runs of one particular kind of change, which would occasionally be produced by chance,” but notes that this was preferable to the subject knowing there would be no such patterns.³³

To most people, the idea of something being “random” means it is unpredictable; but to Peirce, random meant that “in the long run any one individual of the whole lot would get taken as often as any other.”³⁴ When smaller samples from a larger class are drawn in this way, Peirce said that the smaller group would show the same characteristics of the larger group. He called this the “rule of induction.”³⁵ Chapter 7 will delve into random assign-

ment in the social sciences; today, we more typically think of randomly assigning people to different treatments, but, as Peirce’s and others’ early work showed, it is important to randomize as much as you can, including the stimuli in an experiment.

As an aside, to illustrate the scope of Peirce’s abilities, he is also known for developing a theory of semiotics, signs and symbols—one of the classic theories still in use today in critical/cultural work.³⁶

RONALD FISHER'S PLOTS AND TEA

Decades later, Ronald Aylmer Fisher revived and popularized the idea of random assignment of treatments, leading to its widespread use.³⁷ Fisher was an agricultural scientist outside London in the 1920s and ’30s. He is credited with inventing the design of experiments, including many of the statistics, concepts, and procedures still in use today.³⁸ For example, Fisher developed **analysis of variance**, whose statistic, the **F test**, was named after him, and also the idea of **Latin Square**, a procedure for ordering treatments to

compensate for systematic error or control unintended variation instead of randomizing (more about this in chapter 7).ⁱ Fisher was also credited with proposing the probability values of .05 and .01 for statistical significance, which is a scientist's way of judging whether something deviates from chance.³⁹ A *p* value of .05 means there is a “one in twenty chance that the result is mistaken.”⁴⁰

Fisher's job was to test various fertilizers on crops at the Rothamsted Agricultural Experimental Station in England. Because of this, many of the terms still used in experiments today have an agricultural basis, such as **split plot designs**, which actually comes from the plots in fields that were split into separate sections so that each could receive a different treatment. When Fisher arrived at the station, it had been standard to test a different fertilizer every year. But Fisher realized that each year brought inherent differences in rainfall, temperature, weed growth, drainage, and other

factors.⁴¹ He coined the term *confound* when he realized there was no way to separate the effects of the fertilizers from these other conditions that were out of his control.⁴² To determine whether the outcome was due to the fertilizer or something else, he decided to include all of the treatments—in his case, the many different kinds of fertilizers—in the same experiment. By cutting up the fields into small plots, with the pieces divided up into different rows and each row given a different treatment, he reasoned that the yearly differences would apply to all the treatments, effectively “controlling” those confounding conditions.

In addition, Fisher also is credited with popularizing random assignment.⁴³ This also grew out of his efforts to control the various conditions when he realized that systematically assigning different fertilizers to the fields could not rule out confounding factors associated with the soil and fields themselves. To solve this, he randomly assigned the fields to receive different treatments. In social science today, we think of random assignment as applying to the subjects who are assigned to different treatments; for Fisher, the fields were



Ronald Fisher

Wikimedia Commons

ⁱAs a geneticist, Fisher was also interested in evolution and eugenics—the idea that selective breeding could improve the human race. Eugenics had been discussed since early Greece and Rome, and was a respectable scientific topic in his day, supported by many including George Bernard Shaw and Alexander Graham Bell. Fisher helped found a eugenics society at Cambridge University. The idea was thoroughly discredited after being associated with the genocide policies of Nazi Germany. (Gregory Cochran and Henry Harpending, *The 10,000 Year Explosion: How Civilization Accelerated Human Evolution* [New York: Basic Books, 2009]; Encyclopedia Britannica, Sir Ronald Aylmer Fisher, <https://www.britannica.com/biography/Ronald-Aylmer-Fisher>, accessed March 30, 2018; Famous Scientists, <https://www.famousScientists.org/ronald-fisher/>, accessed March 15, 2018.)



Vanderdecken, Wikimedia Commons

his subjects. In the sense of random assignment, “random” means by a chance procedure, such as flipping a coin, and ensures that each participant has an equal chance of being assigned to any of the treatment or control groups. This helps ensure that any systematic differences are equally distributed across the different groups, so that any differences can be attributed to the treatment, not something inherently different about the people in the groups. Chapter 7 will discuss random assignment and how it is achieved in more depth.

One of Fisher’s more relatable experiments that used random assignment is the now-famous Lady Tasting Tea experiment. It also illustrates the idea of a **null hypothesis**, or the supposition that there will be no difference between those who get the treatment and those who do not. As the story goes, a woman who claimed she could tell if the milk was added before or after the tea was given four cups where the tea was added first, and four cups where the milk was added first, the order of which was randomized. The null hypothesis was that she could *not* tell them apart. The chance of someone guessing all eight correctly was one in seventy; the woman in the study purportedly got all eight correct.⁴⁴ The tea experiment was supposedly a summer afternoon of fun, not a scientific investigation with published results. Fisher described it in chapter 2 of his book *The Design of Experiments*.⁴⁵ Nowhere does he say if it was actually conducted or give the results, so much of this is folklore.⁴⁶ Experiments had been done for hundreds of years before Fisher, but they were idiosyncratic, varying with each experimenter. Fisher’s was the first book that systematically codified how to do experiments.

B. F. SKINNER: SMALL SAMPLES, HIGH TECH

Fisher’s principles were just becoming popular when B. F. (short for Burrhus Frederic, but friends called him Fred⁴⁷) Skinner started studying the behavior of rats.⁴⁸ Skinner is popularly known for creating the Skinner box, a device he used to train rats to push a lever for food or to stop electric shocks that he used in developing his theory of operant conditioning.ⁱⁱ Not to belabor the details of his theory or studies here, the premise was

ⁱⁱHe developed another box for pigeons and one for babies; a student developed one for dogs, and much later, another student inspired by Skinner developed a human Skinner box—the study carrels of today. See Rutherford 2007, 2003.^{56, 57}

that animals learned to respond to stimuli, and that reinforcing a behavior makes the animal repeat it. Although Skinner's ideas were sometimes controversial and misunderstood, his work and theory continues to inspire research today, including in human psychology, animal behavior, education, marketing, health, social work, sports training, and others.⁴⁹ The discussion here is devoted to his ideas about experimental research methodology, as that is the purpose of this book.

Skinner is credited with developing a kind of experimental research in psychology that he called the *experimental analysis of behavior*.⁵⁰ It was in contrast to the deductive approach used in most psychological experiments up to that point, which first formed a hypothesis and then tested to see if it could be falsified—the approach still most often used today. Skinner's approach was **inductive** and data-driven, characterized by observing and empirically measuring behavior. He never worked with formal hypotheses, saying, “If I engaged in Experimental Design at all, it was simply to complete or extend some evidence of order already observed.”⁵¹ He published his methods in his book *The Behavior of Organisms*⁵² in greater detail than normal for the time.⁵³

Most of his research was done with individual cases, running one subject at a time or in small groups; four rats were used in most studies in *The Behavior of Organisms*.⁵⁴ He defended his use of small samples and described when they might be better than large samples⁵⁵—for example, when studying populations where recruiting a large number of homogeneous subjects is difficult, such as special education students, mentally ill patients, or the disabled.⁵⁸ Today, this approach is called *single-case experimental design* and is still in use for such populations, especially in clinical work.⁵⁹

One of the reasons why it was difficult for Skinner to study large groups was the constraint of needing a Skinner box for each subject. He only had four such boxes (which he preferred to call “operant chambers”) in his lab.⁶⁰ He describes how he once got a grant to build enough boxes to measure 24 rats at the same time and aggregate the data in a single, mean performance curve. He and his colleague were able to run 95 rat subjects in one study, and Skinner wrote that “the possibility of using large groups of animals greatly improves upon the method as previously reported, since tests of significance are provided for and properties of behavior not apparent in single cases may be more easily detected.”⁶¹ But most often, he ran individual subjects or samples that were too small to perform any



B. F. Skinner

Msanders nL Wikimedia Commons

statistical analysis. He relied on a functional analysis to determine if the intervention worked. For example, when six out of eight pigeons performed the behavior they were trained to perform, “the resulting responses were so clearly defined that two observers could agree perfectly in counting instances.”⁶² He felt statistics were overemphasized when what was really needed were more rigorous controls and techniques.⁶³ He credits Pavlov with the insight “control your conditions and you will see order.”⁶⁴

Skinner also moved experimental methods beyond researchers’ observations to more objective measuring instruments. Skinner had been an inventor since he was a boy, when he built a device to remind him to hang up his pajamas.⁶⁵ To work in conjunction with the Skinner box, he devised a “cumulative recorder” that used a pen and roll of paper to automatically record when a rat pressed a lever or pigeon pecked a key. Use of response rates was widely adopted by other researchers.⁶⁶ He went on to design what he called a “gadget” or “apparatus” for many other uses.⁶⁷ For example, the kymograph used a stopwatch to record the movement and sounds of rats running around.⁶⁸ Another device used a string on a spindle to record data in a curve. “I knew that science made great use of curves, although, so far as I could discover, very little of pips on a polygram,”⁶⁹ he wrote of his previous measures. “As it turned out the curve revealed things in the rate of responding, and in changes in that rate, which would certainly otherwise have been missed.”⁷⁰ Response curves continue to be an important principle in many disciplines.⁷¹ Another invention was his adaptation of a pharmacist’s pill-making machine to make uniform rat food pellets, overcoming the problem of uncontrolled variation in the size of existing rat food. When he realized how much rat chow he would need to make to reinforce every behavior, he started reinforcing only once every minute, which set off a new program of research into “periodic reinforcement.”⁷²

These kinds of practical problems drove much of his research, and he came to see the problems as serendipitous rather than annoying. For example, when the rat pellet delivery device jammed, he wrote, “At first I treated this as a defect and hastened to remedy” it.⁷³ Then, he realized what looked like a setback actually afforded the opportunity to develop “extinction curves.” “It is still no exaggeration to say that some of the most interesting and surprising results have turned up first because of similar accidents.”⁷⁴

Skinner calls science a “disorderly and accidental process” without “a well-defined beginning and end.”⁷⁵ *The Behavior of Organisms*⁷⁶ illustrates the messiness of science and the role of serendipity, or finding one thing while looking for another. (For a contemporary example, see Study Spotlight 2.2 on “Messy” Research.) In Skinner’s case, he was open to seeing it as such. In his own words: “Here was a first principle not formally recognized by scientific methodologists: When you run onto something interesting, drop everything else and study it.”⁷⁷

This is still good advice today.

STUDY SPOTLIGHT 2.2

An Example of “Messy” Research

Babad, E., E. Peer, and Y. Benayoun. 2012. “Can Multiple Biases Occur in a Single Situation? Evidence From Media Bias Research.” *Journal of Applied Social Psychology* 42 (6): 1486–1504.

This study is a contemporary example of the messiness of research. Such candid explanations of how research veered off course and produced unintended results and counterintuitive findings as found in this article are rare in journals. Yet it is more frequent than it would appear, and it is refreshing and reassuring to find such a frank discussion in a paper. This article also illustrates how serendipitous results can arise from studies that do not go exactly as planned.

The first sentence of this paper acknowledges that the idea emerged because of unintended outcomes of experimental research. The study concerns the cognitive and emotional biases that influence people’s judgments and behavior. In these researchers’ domain, that meant the unintentionally biased nonverbal (NV) behavior of television broadcasters when they interview political candidates. The problem these researchers studied arose as a function of some of the best techniques in experimental design—specifically, designing experiments that are free of interfering influences is unlike what occurs in real life. Furthermore, experimentalists’ tendency to study one bias at a time is unlike what occurs in the complexity of daily life, where multiple biases operating at the same time is more likely. The authors concede that the idea did not occur to them initially, nor did published studies point to the possibility of multiple biases. They say, “We must admit that initially we did not conceive the notion of multiple biases, and the phenomenon occurred only in the framework of conventional single-bias research on media bias in interviewers’ NV behavior. This was unsettling at first, because the unplanned bias seemed to interfere with the effect magnitudes of the investigated media bias.”⁷⁸

The researchers undertook the arduous task of conducting replications of their own study with seven different samples before concluding that two different biases were at work. This paper reports the meta-analyses of them. At the conclusion of the studies, they frankly say where findings were not expected, counterintuitive, surprising, and disappointing. They admit when findings are difficult to explain. For example, “The halo effect was not really expected in this research because the experiment was not designed to create halo effects. . . . The most unexpected and counterintuitive result, perhaps with the most far-reaching conceptual and applied implications, was the independence of the two bias phenomena, as indicated by rejection of the vicarious halo effect hypothesis . . . It is not easy to explain this effect, and it is, indeed, counterintuitive.”⁷⁹

Rather than attempting to cover up and posture, these researchers reveal the struggles and messiness of experimental research that is more common than published articles lead researchers to believe.

STANLEY MILGRAM SHOCKS THE WORLD

Stanley Milgram’s studies are some of the most recognizable, known to scholars and lay people alike. Colloquially known as the “shock studies,” Stanley Milgram’s experiments on obedience were designed to discover the causes behind why people obey authority figures even when what they are asked goes against their own conscience. In the 1960s,

the Nazi war criminals trials were being held, so interest in this question was intense. Milgram's experiments are known for the criticism they incurred from psychologists for causing psychological harm to those who participated,⁸⁰ and their deceptiveness. (For more on deception, see More About box 2.3 on Deception.) This criticism ushered in a larger discussion about ethics in social science research. These studies also are credited with helping create IRBs to ensure research did not harm the people who volunteered for them (see more about IRBs in chapter 11).

While these studies are mainly remembered as cautionary tales about ethical research, there are other aspects of the studies that lend themselves to lessons in experimental design. The main structure of all Milgram's studies was for a **naïve subject**—someone who did not know what the study was about—to “teach” another person a set of word pairs. The subject was told that the goal was to see if punishment, in the form of electric shocks, improved learning. In this case, the “learner” was an accomplice, or **confederate**, of the experimenter who knew the purpose of the study and was only pretending to be shocked. Experiments that used confederates and deception such as this were popular around this time. Milgram used a **debriefing** session at the end of each experiment, telling subjects the true purpose of the experiment, that they had not actually harmed anyone, and that their responses were normal.⁸¹ After two articles were published and criticism had begun to appear, Milgram began reporting the steps he took to minimize harm to subjects and the poststudy surveys he conducted to understand if harm had occurred.⁸² Diana Baumrind dismissed the postexperiment self-reports of Milgram that showed more than 80% of subjects of their deceptive experiments said they were glad to have participated,⁸³ saying they were “tacked on as an afterthought,” and “After all, if self-reports could be regarded as accurate measures of the impact of experimental conditions, we could dispense entirely with experimental manipulation and behavioral measures.”⁸⁴

Confederates are out of vogue today; however, debriefings are still a best practice when experiments involve deception, although studies such as Milgram's would not likely be approved by an IRB today. Notably, however, the popularity of “experiments” in this genre continues on television shows such as ABC's hidden camera series *What Would You Do?* No IRB approval is necessary when journalists or entertainment producers are doing it.

Chapter 1 explained the concepts of cause and effect. The effect that Milgram was interested in was obedience as measured by the dependent variable of the maximum shock that a subject would give to another person, which ranged from 0 to 30.¹²⁴ The causes he was examining were a myriad of independent variables including how close the subject was to his or her victim, if the subject could hear protests and cries, had to touch the victim, the authority of the person giving the instructions to administer shocks, and more.

MORE ABOUT . . . BOX 2.3

Deception

While deception in social science experiments may no longer be as dramatic as Stanley Milgram's shock experiments, neither is it a thing of the past. Its use varies among disciplines; it is common in sociology and social psychology within the parameters of the major associations⁸⁵ but proscribed in economics,⁸⁶ although withholding information is not considered deceptive, so that is allowed.⁸⁷

Deception is defined as intentionally providing false information and withholding information in order to mislead.⁸⁸ This does not prevent researchers from withholding the hypotheses, conditions of the experiment, or other aspects of the research that would cause subjects to change their behavior.⁸⁹ However, withholding such information crosses the line if the nature of it would cause subjects to not agree to participate.⁹⁰

Deception is prohibited in studies that may cause physical pain or severe emotional distress.⁹¹ Guidelines mandate researchers not use deception unless it is justified by the study's value or there is no other feasible way.⁹² However, as Baumrind says, "It takes little to convince a researcher or a review board of his or her peers that the long-range benefits of a clever bit of deceptive manipulation outweigh the short-range costs to participants of being deceived."⁹³

Deceptive Research

Nor does the 20/20 vision of hindsight always clear things up. After his prison experiment, Philip Zimbardo wrote about considering other methods than the ones he used and employing an objective observer to monitor the study.⁹⁴ Yet in years immediately after, Zimbardo and colleagues used hypnosis to induce partial deafness that resulted in subjects experiencing paranoia, misinforming subjects about the purpose and what they would undergo.⁹⁵ Just like the prison experiment, this and another study that involved misleading subjects⁹⁶ received IRB approval. In 1977, Stanley Milgram's shock study was replicated with children as young as six.⁹⁷ This study was conducted in a country without an IRB but was approved by the department chair and other faculty, as well as the children's teachers and principals. The study reports the children who participated exhibited "loud nervous laughter, lip biting, trembling."⁹⁸ At the end of the article, the authors say, "it is indeed surprising to find that relatively few social psychologists have followed up on Milgram's pioneering work on obedience."⁹⁹ This study was published in one of the top social psychology journals then and now, which also published one of Milgram's.¹⁰⁰ This is not meant to imply that these researchers are unethical or even tone deaf, but that detecting issues with deception can be harder than one thinks.

Content analyses show the trajectory of deceptive research. Beginning in 1921, there was rarely any deception through the 1930s in the leading social psychology journal, *Journal of Personality and Social Psychology*. Its use grew gradually until the 1950s and then grew significantly between the 1950s and 1970s with the rise of experimental methods.¹⁰¹ There was a decrease from the 1980s to 1994 when this study ended.¹⁰² It found that computers and other bogus devices partially replaced confederates to mislead subjects. Misleading consent and false feedback to subjects fell after 1969 but rose again in 1992, and changes in topics studied explained declines in deception between 1978 and 1986.¹⁰³ Deceptive practices were found in as many as 66% of the articles examined.¹⁰⁴ Other

(Continued)

(Continued)

studies show similar results. Between 1971 and 1974, 54% of psychology experiments used deception,¹⁰⁵ up from 18% in 1948 and 37% in 1963.¹⁰⁶ More recent figures find deception in about 33% of social psychology studies.¹⁰⁷

That deception can harm subjects is obvious. In addition to physical pain or psychological distress, deception harms study subjects in that it undermines their autonomy and violates their right to choose to participate.¹⁰⁸

But deceptive research also has the potential to harm more than study subjects, with effects extending to the entire research profession. As many social scientists have noted, voluntary participation in research studies has been declining. This is attributable in no small part to public suspicion of research based on its history of harm, deception, and abuse of power.¹⁰⁹ It is not only the ability to recruit subjects that suffers, but public trust and confidence in the findings of studies can be eroded, as well as the reputation of the entire academy. Society becomes less willing to support research.¹¹⁰ Furthermore, the willingness of a few experimentalists to use deception has the ability to induce others to do so as well,¹¹¹ thereby undermining the moral agency of the researchers themselves.¹¹² Detriments to society at large include an overall loss of trust and increasing suspicion as the social contract to tell the truth and keep promises is violated.¹¹³

Using Deception Ethically

The most common reason for the use of deception includes the need to control demand characteristics, or the tendency for subjects to alter their behavior to fit with the goals of the study.¹¹⁴ Baumrind contends that using deception promotes the very thing it was designed to prevent: because people know deception is routinely used in experiments, subjects are no longer naïve—that is, they do not believe the study is real.¹¹⁵ At the same time, other research finds that deception does not affect the external validity of studies.¹¹⁶

The most common techniques for ethically using deception, other than avoiding it completely, include:

- Giving subjects informed consent that includes the true purpose of the study without giving away hypotheses, and describes the procedures and what they will do.¹¹⁷
- If any information is withheld, debriefing subjects immediately after the study is over.¹¹⁸ Also, subjects should be asked if they agree to postpone receiving all the information until the debriefing.¹¹⁹ Recognize, however, that debriefing does not undo harm.¹²⁰
- Allow subjects to withdraw from the study at any time, including withdrawing their data after they have been debriefed.¹²¹

Experimentalists should realize they might be inclined to choose deceptive methods because they are easier than nondeceptive ones. Instead, researchers should consider nondeceptive methods such as natural experiments where no manipulation is involved or unobtrusive instruments where the subject is unable to alter his or her responses, such as psychological measures of heart rate, or latency response, which measures how long a person takes to respond to a stimulus. Natural experiments are covered in chapter 4, and unobtrusive instruments in chapter 10. There is more about ethics and

deception in the context of IRBs in chapter 11. Other suggestions that have been made for nondeceptive experiments include having researchers be their subjects and examine their motives and behavior,¹²² or using “surrogate subjects”—that is, telling people about the purposes and procedures to see if they find them acceptable.¹²³

Finally, my own advice is to apply the golden rule—that is, treat others as you would like to be treated. Put yourself in the shoes of your research subjects and ask if you would have any objections to being tricked, lied to, or learning something painful about yourself. Consider the consequences, including what it would be like to have your employer, family, or friends find out about things you would prefer to keep private.

As Milgram writes, “The crux of the study is to systematically vary the factors believed to alter the degree of obedience to experimental commands.”¹²⁵ In all, he conducted eighteen different experiments involving individuals in original journal articles,¹²⁶ and the entire series was reported in his book *Obedience to Authority*.¹²⁷

Chapter 1 also introduced the “so what” question, or explaining why an experiment is important. Milgram answered the “so what” question by citing practical problems his research could shed light on, from war crimes to workers obeying their bosses. He also devised a clever way to answer the criticism of research that only confirms what we already know rather than studies that reveal something new or unexpected. He built in a procedure for determining if his were obvious and intuitive findings by conducting a separate study where he invited subjects to a lecture on the topic of his study and described the experiment without disclosing the results.¹²⁸ After the lecture, subjects were asked several questions about what he or she would do and what they thought others would do. All 110 subjects said they would refuse to administer shocks at some point, and that virtually all the other subjects would also. That was the opposite of what had happened in the studies, with 65% of the subjects going all the way to the highest level of shocks.¹²⁹



Stanley Milgram

Thus, Milgram was able to show that his results were not intuitive and, in fact, were the reverse of what people expected would happen.

Just as Lind did with his scurvy experiments, Milgram employed a combination of theory and observation to test the causes of obedience. For example, he cited conflicting theories about women as the reason for experiment 8, where he used female subjects for the first time. Anachronistic as it is today, some theory and evidence at that time pointed to women being more obedient and less aggressive than men, which should lead them to give the shocks *more* frequently than men. Conflicting theory and evidence showed women were more empathetic than men, which predicted they would give shocks *less* frequently than men.¹³⁰ In fact, women and men performed about the same on the shock studies, but they talked about their reasons for administering the shocks in different terms in the postexperiment interviews. Chapter 3 will cover the importance of theory to experiments and how observation and experience also can drive experimental questions.

Milgram also addresses the topic of sampling and participants in his studies. He explicitly rejected the use of undergraduate students, which is fairly common today although heavily debated.¹³¹ Noting that using students would have been far easier for him, Milgram gave two reasons that he thought might minimize the findings because they could have heard about the studies from others who had already participated, and were too homogeneous, being of similar age, intelligence, and familiar with psychological studies.ⁱⁱⁱ Instead, he wanted a “wide range of individuals drawn from a broad spectrum of class backgrounds.”¹³² So he ran a newspaper ad to recruit participants from the surrounding community, luring them with a \$4.50 cash incentive. Milgram writes about using forty subjects for each condition in all the studies but does not mention how he arrived at that number. Today, a **power analysis**, a statistical technique to determine how many subjects a particular study needs, would be used. That will be covered in chapter 8 of this book. Nor does he randomly assign subjects to condition, a feature designed to ensure equivalence that had already been discovered by others but used to order treatments rather than to assign subjects. Instead, Milgram used a matching technique similar to that employed by Lind in the scurvy studies. He writes about how it was important to balance subjects in each condition by age and occupation, so he divided them into categories of skilled and unskilled workers

ⁱⁱⁱAs to the question of using students or not, after the adult citizens had been used in all the experiments, Milgram examined their results in relation to the students he had used in pilot studies and found no differences. Milgram, *Obedience to Authority: An Experimental View*, 170.

versus white collar and professionals, and also into three age groups, and assigned them based on those characteristics.¹³³

The many treatment conditions, eighteen in all, were carried out one after the other, moving down to “more moderate alterations of the situation.”¹³⁴ For example, in experiment 7, he started with the experimenter being physically close to the subject and gradually increased the distance between them until the experimenter was out of the room entirely, giving orders via telephone. The same thing happened in experiments 1 through 4, when the subject’s proximity to the victim was systematically altered, bringing the subject and victim increasingly closer. This illustrates the principle of experiments maximizing variation by starting with the strongest manipulation first and if effects are found, moving to lower levels of the manipulation until the effect disappears. This will be covered more in chapter 9. In all the studies, Milgram also controlled for confounds by using recorded protests of the victim so they were the same for every subject and every shock level.¹³⁵

Milgram’s studies employed a variety of measurements, the subject of chapter 10. His main dependent variable was behavioral—the level of “shock” subjects administered, ranging from zero to the highest level of 30, representing a whopping 450 volts. In reality, the lowest level anyone could administer was 20 volts, and Milgram also measured latency of the shocks—that is, how long it took subjects to press the switch and how long subjects held down the switch.

Additional measures included **self-reports** from the subjects via **questionnaire** after the experiment about the level of conflict, tension, and nervousness they experienced during the study, and their estimation of the amount of pain felt by the victim. Milgram also used a variety of **response choices**, including **open-ended questions**, **projective tests**, and **attitude scales**.¹³⁷ All of these will be covered in detail in chapter 10. In Milgram’s first four experiments, subjects also were asked how much responsibility they assigned to the experimenter, subject, and victim for a person receiving shocks against his or her will.¹³⁸ The questionnaire asked the usual demographic variables of political party, religious affiliation, education, and, in this case, length of military service, as that was common in the 1960s and deemed important to how well one followed orders.¹³⁸

In addition, Milgram recorded the conversations between the subjects and experimenter during the study, and the researchers noted their own observations. These **qualitative measures** were used in addition to individual interviews and group discussions after the experiments, which he reported to give context to the findings, an important feature of triangulation discussed in chapter 1.

We also find in Milgram's work examples of pilot studies, one of the subjects of chapter 11, and manipulation checks, the topic of chapter 9. **Pilot studies** are used to test-drive an experiment before the real study is launched in order to work out the kinks. In his pilot studies using undergraduate students, Milgram found subjects needed practice reading the words to the learners, so he incorporated a training session where subjects read ten word pairs before the actual study began.¹⁴⁰ Another important lesson from the pilot studies was that of "vocal feedback."¹⁴¹ Initially, Milgram had the subject shock the victim in another room with the victim remaining silent. He found no variation in how far subjects would go in delivering shocks—everyone went all the way. With no variation in the outcome—no one disobeyed—there was no way to determine the causal mechanisms of disobedience. So Milgram introduced cries and protests from the victim, and pounding on the walls, which caused subjects to stop the shocks at different points, thereby giving him the variation discussed in chapter 1.¹⁴²

One subject of chapter 9 is how to determine whether manipulations are perceived as realistic. For example, if the experiment involves subjects reading messages written in vivid, descriptive language or in nonvivid, abstract style, it is necessary to know if subjects who got the vivid writing perceived it was more vivid than those who got the nonvivid

versions.¹⁴³ For this, a **manipulation check** is used. In Milgram's case, he wanted to know if subjects believed the victims actually received painful shocks and if they believed they were the ones administering them. Only two out of the forty subjects answered those questions with "no" on the postexperiment questionnaire, so the manipulation was deemed to have worked.

PHILIP ZIMBARDO: RAISING CONSCIENCES IN A STANFORD BASEMENT

Shortly after Stanley Milgram's shock studies, Philip Zimbardo conducted another noteworthy study in the same vein. On a personal note, Milgram and Zimbardo had met in high school in 1949 in the Bronx.¹⁴⁴ They reconnected in 1960 as assistant professors, with Milgram at Yale and Zimbardo at New York University. Coincidentally, Zimbardo had originally constructed



Philip Zimbardo

the basement lab at Yale that Milgram used for his shock studies after he moved out of the “elegant interaction lab.”¹⁴⁵ But it was a different basement across the country that was to go down in history for, among other things, the ethical questions it raised.

In the summer of 1971, Zimbardo constructed a mock prison in the basement of the Stanford University psychology building for a study funded by the Office of Naval Research, which was interested in the causes of conflict between prisoners and military guards. It was here that twenty-one college students from across the country spent six days playing the roles of either prisoners or guards. But unlike in 1961 when Milgram did his work, in the 1970s, IRBs existed. And Zimbardo and his graduate students did, in fact, get IRB approval. Zimbardo’s book *The Lucifer Effect*¹⁴⁶ describes the steps he took to avoid harming subjects, including prohibiting physical violence. All the participants were given **informed consent**, which included the information that they were free to leave the experiment at any time. Few did. No deception was involved, unlike in Milgram’s study. However, the researchers gave the guards suggestions on how they should behave rather than allowing all the behaviors to develop naturally. This study has gone down in history as one of the most harmful ever. For example, it used psychological techniques to induce boredom and take away subjects’ privacy, their sense of individuation, and power. The subjects’ reactions to the study were so intense, with the students playing both guards and prisoners internalizing their new identities to an alarming degree, that the experiment was stopped after six days instead of the planned two weeks because of the psychological and emotional trauma.¹⁴⁷ Some prisoners went into screaming, out-of-control rages and hunger strikes, while some guards exhibited sadistic behavior and even attacked prisoners with fire extinguishers. Anticipating concerns, the team did extensive debriefing of each subject immediately after the study concluded and followed each subject for a year looking for ill effects.¹⁴⁸ In his response to critics, Zimbardo pointed out that the only request received by the American Psychological Association to investigate the study had come from him,¹⁴⁹ and the research was fully cleared. A proposal for a follow-up study was turned down by the IRB, and Zimbardo apologized to the participants in the 2007 book.¹⁵⁰

The prison study was a laboratory experiment, normally criticized for being highly artificial, so Zimbardo and colleagues took great pains to make it as realistic as possible, just as Milgram had. For example, they talked the Palo Alto police department into having real officers “arrest” the subjects.^{iv} They outfitted the subjects who were guards in clothing similar to real prison uniforms. Inmates stayed in cells with steel bars 24 hours a day,

^{iv}The researchers had to be creative to get the police to cooperate, using the enticement of a TV station putting it on the evening news as good publicity for the force, according to Zimbardo in *The Lucifer Effect*.

while the guards worked regular shifts. They staged visiting days with real friends and relatives, and held mock parole hearings.¹⁵¹ It is hailed as a prime example of realism to this day.¹⁵² The researchers noted that while they wanted their experiment to be as real as possible, conducting it in an actual prison was not an option because “There are too many uncontrolled variables in the real world, or in the ‘field,’ as social scientists call it. That’s the comfort of laboratory research: The experimenter is in charge.”¹⁵³ They point out that many studies had already been done in existing penal institutions, which they called **natural experiments**.¹⁵⁴ The differences between laboratory experiments, natural, and **field experiments** will be explored in chapter 4. Zimbardo explained, “There have been studies of actual prison life by sociologists and criminologists, but they suffer from some serious drawbacks. Those researchers are never free to observe all phases of prison life . . . They can see only what they are allowed to see.”¹⁵⁵ They explained their choice of a realistic-as-possible lab experiment by noting that their research question could not be studied using real prisons and prisoners because it was necessary to separate the effects of the prison environment from the characteristics of its inhabitants—that is, there were confounds, described in chapter 1, in the real world. In this case, the study was undertaken because of the belief that the violence and cruelty found in prisons was due to the antisocial personalities of inmates and authoritarian characteristics of guards. Thus, they decided to create a new solution using a group of undifferentiated people who did not already possess the antisocial and sadistic personalities presumed responsible for prison conditions. The mock prison was “entirely populated by individuals who are undifferentiated in all essential dimensions from the rest of society.”¹⁵⁶ Thus, by holding constant the situational aspects assumed to be responsible for prison conditions, the study showed that behaviors in prisons could be reliably attributed to the situation rather than just the personalities of the people in them.

To get their “normal” subjects, the research team did a combination of Milgram’s purposeful selection based on specific criteria and Fisher’s random assignment. They recruited subjects using a newspaper ad promising \$15 a day for participation. Unlike Milgram, who thought college students would bias the study, Zimbardo chose students on purpose because they were so similar, or **homogeneous**. The seventy-five students who responded were not just from Stanford but colleges across the country that happened to be in Palo Alto for the summer. The recruits had to answer questions about their families, physical and mental health, and involvement in crime, and also be interviewed by the researchers. The researchers chose the most stable and mature subjects.¹⁵⁷ As with Milgram’s and most other studies of the time, the subjects were male. And even though they were looking for students who were similar to the population of the United States, all but one were White.¹⁵⁸

Following this careful, purposive selection procedure, the chosen subjects were then randomly assigned to be either prisoner or guard. The purpose of the random assignment to condition was to ensure that the subjects who were playing guards were not different from those playing inmates on any important characteristics. The team checked this by giving subjects a battery of psychological tests, which showed no significant differences between either group. Random assignment had worked.¹⁵⁹ Chapter 7 will go into more detail about how this is done.

Just as Milgram had conducted pilot studies, Zimbardo and colleagues went into this study with prior research informing it. The idea arose from a project that a student in one of Zimbardo's classes had done. Zimbardo followed that up with a "field experiment"¹⁶⁰ where he conducted a candid camera-type of study, putting abandoned cars in Palo Alto, CA, and Bronx, NY, and recording people vandalizing them. Ordinary citizens in the blue-collar, working-class Bronx and upper-class, white-collar Palo Alto, all White and well dressed, some encouraging their children, vandalized the cars. From these pilot studies, Zimbardo concluded that conditions that make people feel anonymous could foster antisocial behavior.¹⁶¹ Chapter 11 will discuss how to conduct pilot studies.

The original research article described the prison experimental design as "relatively simple" with "a single treatment variable, the random assignment to either 'guard' or 'prisoner' condition."¹⁶² It was clearly more complicated than Milgram's, where there was only one condition or level of **factor** per experiment, but much less complex than typical experimental designs today. The types of designs will be discussed in more detail in chapter 4 and factorial designs in chapter 6. The study did not even have a formal hypothesis, which the researchers explained by the "exploratory nature" of the study.¹⁶³ They do report multiple statistical tests to support the one general hypothesis offered: "That assignment to the treatment of 'guard' or 'prisoner' would result in significantly different reactions on behavioural measures of interaction, emotional measures of mood state and pathology, attitudes toward self, as well as other indices of coping and adaptation to this novel situation."¹⁶⁴ Milgram had not offered a hypothesis in his initial paper either.

Like Milgram's, the prison study used a plethora of measures, including video and audio recordings of the subjects' actual behavior, researchers' observations, questionnaires using self-reports, and interviews with the subjects.¹⁶⁵ The questionnaires included mood inventories, personality tests, scales measuring authoritarianism and Machiavellianism, and a personality scale that included trustworthiness, orderliness, conformity, activity, stability, extroversion, masculinity, and empathy.¹⁶⁶ Chapter 10 will go into more detail about constructing measures such as these. The researchers also triangulated their data using a combination of quantitative and qualitative techniques. For example, they used an inductive approach to categorize the data from the video and audiotapes and observations, which they converted into scores

on which they performed statistical tests. They also reported findings qualitatively, including quotes from subjects. One strength of the findings was the “consistency in the pattern of relationships which emerge across a wide range of measuring instruments and different observers.”¹⁶⁷ The findings in this paper remain a good example today of how to write this section of an experimental paper. A great deal of space is devoted to the discussion of the unique findings from the study, their meaning, and the reasons behind them. They are related to other research without the section looking like another literature review.¹⁶⁸ The discussion also includes limitations of the study, including the small sample size ($N = 21$), that the conditions in the simulated prison were minimal compared to real prisons, and **demand characteristics**, in this case, the subjects knew they were being observed and wanted to please the researchers. The paper shows how researchers overcame some of these limitations with the data they had—for example, analysis of times when the subjects did not know they were being observed but their cells were bugged. Being able to use data to show how a study’s limitations do not invalidate the conclusions is a useful skill.

Finally, one of the most important things this study does is answer the “so what” question in a comprehensive and definitive way. It was hailed by the American Psychological Association as an exemplar of how psychological research could be applied to solve real-world problems and understood by nonacademics, making research relevant to ordinary citizens.¹⁶⁹ In response to criticisms of the unethical nature of the research, Zimbardo reported on the impact to society and listed all the ways the public had shown an interest in the work, including media stories, calls, letters, requests to speak, and changes to law and policy, among others.¹⁷⁰ While Zimbardo was responding to charges of unethical treatment, all experiments should present a strong rationale for the ability of the research to improve conditions in society and solve real-world problems.¹⁷¹

CONCLUSION

All these famous experiments from the past have some things in common, including creativity. These social scientists were willing to go to great lengths to ensure their experiments found something real rather than just support a hypothesis. Philip Zimbardo enlisted the help of real police officers. B. F. Skinner invented his own measuring devices. In their quest, some ignored their duty to protect subjects, leading today’s researchers to better understand the trade-offs between harm and reality. Experimentalists should still strive to make their study conditions as realistic as possible; if a study will test TV messages, then subjects should view them via TV, not in written script format, for example. But preventing harm is paramount. Chapter 11 will give more specific advice for avoiding or minimizing deception in experiments, along with other ethical issues.

Nor were these scientists afraid to try something new. Important advances in knowledge are made by knowing when to stand on the shoulders of others, and also when to strike out on one's own. Too many are afraid to deviate from what has already been done, or measure anything differently than it has already been measured. This results in studies that confirm others' findings but little else, when the purpose of science is to create new knowledge.

These historical examples also show that research is messy. It does not always proceed as neatly as the write-up in a journal makes it appear. Most of these researchers document the disorderly process in their books rather than journal articles. The books of Skinner, Milgram, and Zimbardo are good examples.

Finally, these studies show it is important not to harm subjects and to deceive them as little as possible and within ethical guidelines (see More About box 2.3 on Deception). It is essential to find valid answers to questions while balancing the duty to protect subjects. Deception should be avoided if at all possible, and when not possible, used as little as necessary. Debriefing participants after the study and offering to send them a report about the findings once concluded helps treat people with dignity and respect. These steps will also help ensure that one's own studies go down in history for all the right reasons.

For Review-No Commercial Use(2023)

Common Mistakes

- Not making studies realistic
- Using deception when it could be avoided
- Replicating others' findings rather than creating new knowledge

Test Your Knowledge

1. In his study on scurvy, James Lind assigned subjects to treatment conditions using _____.
 - a. Random assignment
 - b. A matching strategy
 - c. A representative strategy
 - d. A snowball strategy

2. By some accounts, one of the first experimentalists to use random assignment was _____.
 - a. James Lind
 - b. Isaac Newton
 - c. C. S. Peirce
 - d. Albert Einstein
3. Ronald Fisher's prediction that a lady would not be able to tell if the tea or milk came first is an example of _____.
 - a. An alternative hypothesis
 - b. A null hypothesis
 - c. A duh hypothesis
 - d. A theoretical hypothesis
4. B. F. Skinner was known for using _____.
 - a. Small samples
 - b. Large samples
 - c. Random samples
 - d. Purposive samples
5. When one of B. F. Skinner's devices jammed and wouldn't work, he recognized this as _____.
 - a. A problem to be rectified
 - b. The messiness of science and the role of serendipity
 - c. Sabotage
 - d. A way to control confounds
6. Stanley Milgram used latency measures, self-reports, and how subjects behaved in his obedience studies. This illustrates _____.
 - a. Measurements
 - b. Response choices
 - c. Controlling for confounds
 - d. Maximizing variation
7. Stanley Milgram also used attitude scales. This illustrates _____.
 - a. Measurements
 - b. Response choices
 - c. Controlling for confounds
 - d. Maximizing variation

For Review-No Commercial Use(2023)

8. The purpose of a pilot study is _____.
a. To discover if the manipulations work as expected
b. To discover if subjects think the experiment is real
c. To test-drive the experiment to figure out what needs to be changed
d. To ensure equivalent groups
9. In which of the following studies was deception involved?
a. C. S. Peirce's study on weights
b. Ronald Fisher's lady tasting tea experiment
c. Stanley Milgram's obedience to authority experiments
d. Philip Zimbardo's prison experiment
10. For the prison experiment, Philip Zimbardo used _____.
a. Purposive sampling
b. Random assignment
c. Both a and b
d. None of these

For Review-No Commercial Use(2023)

Answers

- | | | | |
|------|------|------|-------|
| 1. b | 4. a | 7. b | 9. c |
| 2. c | 5. b | 8. c | 10. c |
| 3. b | 6. a | | |

Application Exercises

1. This chapter does not pretend to cover all the creative pioneers of experimental designs in the social sciences. Identify some of the others and do your own research on them—for example, Kurt Lewin, Leon Festinger, Solomon Asch, Ivan Pavlov, Paul Lazarsfeld, Muzafer and Carolyn Sherif, among many others. Especially look for women and scientists of color. Read what others have written about these experimentalists and their work, but also look up and read their *original* research papers or books written by these social scientists themselves.
2. Famous experiments in history seem to have captured the imagination of filmmakers recently. Watch one of these and write two pages on the importance of ethical research:

- *The Experimenter*, about Stanley Milgram's famous electric shock studies, 2015. Available on Netflix, Amazon, and Microsoft Movies and TV. This was a feature film at the Sundance festival in 2015. Milgram also filmed the actual experiments, and you can see real footage, with Milgram himself, on YouTube and other services by searching for "Stanley Milgram experiment video."
- *The Stanford Prison Experiment*, about Philip Zimbardo, 2015. Available on Amazon, Microsoft Movies and TV, and iTunes. Zimbardo has a website with more information on the research: <http://www.prisonexp.org>.

Suggested Readings

Nothing is more enlightening than reading an original study rather than only what others say about it. Thus, I recommend the following:

- Haney, Craig, Curtis Banks, and Philip Zimbardo. 1973. "Interpersonal Dynamics in a Simulated Prison." *International Journal of Criminology and Penology* 1: 69–97.
- Milgram, Stanley. 1964. "Group Pressure and Action Against a Person." *Journal of Abnormal and Social Psychology* 67: 371–382.
- Milgram, Stanley. 1965. "Liberating Effects of Group Pressure." *Journal of Personality and Social Psychology* 1: 127–134.
- Milgram, Stanley. 1965. "Some Conditions of Obedience and Disobedience to Authority." *Human Relations* 18 (1): 57–76.
- Peirce, Charles Sanders, and Joseph Jastrow. 1885. "On Small Differences of Sensation." *Memoirs of the National Academy of Sciences* 3: 75–83.
- Skinner, B. F. 1948. "Superstition in the Pigeon." *Journal of Experimental Psychology* 38: 168–172.

For an in-depth discussion of the pros and cons of creativity in research, read:

Voosen, Paul. 2015. "For Researchers, Risk Is a Vanishing Luxury." *Chronicle of Higher Education*. www.Chronicle.com. This is a premium article, meaning there is a charge or a subscription is required. Check your college or university library for access.

For those especially interested in the history of deception, read this book:

Korn, J. H. 1997. *Illusions of Reality: A History of Deception in Social Psychology*. Albany: State University of New York Press.

For an entertaining read that shows the more human side of many of these experimentalists, read:

Slater, Lauren. 2005. *Opening Skinner's Box: Great Psychological Experiments of the 20th Century*. New York: Norton.

Notes

1. Nancy Felipe Russo and Agnes N. O'Connell, "Models from Our Past: Psychology's Foremothers," *Psychology of Women Quarterly* 5, no. 1 (1980): 11–54.
2. Ibid.
3. Kendra Cherry, "10 Women Who Changed Psychology: A Closer Look at Women in Psychology," *VeryWellMind* <https://www.verywellmind.com/women-who-changed-psychology-2795260>. Accessed March 21, 2018.
4. For more biographies of women in psychology, see the special issue of *Psychology of Women Quarterly*, 1980, vol. 5, issue no. 1.
5. Agnes N. O'Connell and Nancy Felipe Russo, "Models for Achievement: Eminent Women in Psychology," *Psychology of Women Quarterly* 5, no. 1 (1980): 6–10.
6. Laurel Furumoto, "Mary Whiton Calkins (1863–1930)," *Psychology of Women Quarterly* 5, no. 1 (1980): 55–68.
7. Ibid.
8. Ibid.
9. Joseph Jastrow, "A Study in Mental Statistics," *New Review* 5 (1981): 564.
10. Mary Whiton Calkins, "Community of Ideas of Men and Women," *Psychological Review* 3 (1896): 426–430.
11. Cherry, "10 Women Who Changed Psychology."
12. "Psychologist Mamie Phipps Clark Profile: Important Contributor to the Self-Concept Among Minorities Discussion," *VeryWellMind* <https://www.verywellmind.com/mamie-hipps-clark-biography-2796022>. Accessed March 21, 2018.
13. L. Nyman, "Documenting History: An Interview with Kenneth Bancroft Clark," *History of Psychology* 13, no. 1 (2010): 74–88.
14. "Psychologist Mamie Phipps Clark Profile: Important Contributor to the Self-Concept among Minorities Discussion."
15. Stephen N. Butler, "Mamie Katherine Phipps Clark (1917–1983)," *The Encyclopedia of Arkansas History & Culture* <http://www.encyclopediaofarkansas.net/encyclopedia/entry-detail.aspx?entryID=2938> (2016).
16. Ibid.
17. Cherry, "10 Women Who Changed Psychology."
18. M. D. Ainsworth and B. A. Wittig, "Attachment and Exploratory Behavior of One-Year-Olds in a Strange Situation," in *Determinants of Infant Behavior*, ed. B. M. Foss (London: Methuen, 1969), 113–136.
19. U. G. Wartner et al., "Attachment Patterns in South Germany," *Child Development* 65 (1994): 1014–1127.
20. M. E. Lamb, "The Development of Mother-Infant and Father-Infant Attachments in the Second Year of Life," *Developmental Psychology* 13, no. 6 (1977): 637–648.
21. M. Marrone, *Attachment and Interaction* (New York: Jessica Kingsley, 1998); E. C. Melhuish, "A Measure of Love? An Overview of the Assessment of Attachment," *ACPP Review & Newsletter* 15 (1993): 269–275.
22. Russo and O'Connell, "Models from Our Past."
23. U.S. National Library of Medicine, "Scurvy," *Medline Plus* (2016).
24. James Lind, *A Treatise on the Scurvy: In Three Parts. Containing an Inquiry into the Nature, Causes, and Cure, of That Disease. Together with a Critical and Chronological View of What Has Been Published on the Subject* (Edinburgh: James Murray, and Cochran for A. Millar, 1753).
25. Ibid., 191–192.
26. Ibid., 193.
27. Ibid.
28. R. Barker Bausell, *Conducting Meaningful Experiments: 40 Steps to Becoming a Scientist* (Thousand Oaks, CA: Sage, 1994).
29. Lind, *A Treatise on the Scurvy*.
30. Stephen M. Stigler, "Mathematical Statistics in the Early States," *The Annals of Statistics* 6, no. 2 (1978): 239–265.
31. Charles Sanders Peirce and Joseph Jastrow, "On Small Differences of Sensation," *Memoirs of the National Academy of Sciences* 3 (1885): 75–83.
32. Stephen M. Stigler, "A Historical View of Statistical Concepts in Psychology and Educational Research," *American Journal of Education* 101, no. 1 (November 1992): 60–70.
33. Charles Sanders Peirce and Joseph Jastrow, "On Small Differences in Sensation," in *American Contributions to Mathematical Statistics in the Nineteenth Century*, ed. Stephen M. Stigler (New York: Arno Press, 1980/1885), 80.

34. Charles Sanders Peirce, *Essays in the Philosophy of Science* (New York: Liberal Arts Press, 1957), 217.
35. Ibid., 104.
36. Joseph Brent, *Charles Sanders Peirce: A Life*, 2nd ed. (Bloomington and Indianapolis: Indiana University Press, 1998).
37. Stigler, "Mathematical Statistics in the Early States," 249.
38. David Salsburg, *The Lady Tasting Tea: How Statistics Revolutionized Science in the Twentieth Century* (New York: W. H. Freeman, 2001).
39. Ronald A. Fisher, *Statistical Methods for Research Workers* (Edinburgh: Oliver and Boyd, 1925).
40. Diana C. Mutz and Robin Pemantle, "Standards for Experimental Research: Encouraging a Better Understanding of Experimental Methods," *Journal of Experimental Political Science* 2, no. 2 (2016): 192–215.
41. Salsburg, *The Lady Tasting Tea*.
42. Ibid.
43. Stigler, "A Historical View of Statistical."
44. Salsburg, *The Lady Tasting Tea*.
45. Ronald A. Fisher, *The Design of Experiments* (Edinburgh: Oliver and Boyd, 1937).
46. Salsburg, *The Lady Tasting Tea*, 8.
47. Robert P. Hawkins, "The Life and Contributions of Burrhus Frederick Skinner," *Education & Treatment of Children* 13, no. 3 (1990): 258.
48. B. F. Skinner, "A Case History in Scientific Method," *The American Psychologist* 11, no. 5 (1956): 221–233.
49. Joshua B. Plavnick and Summer J. Ferreri, "Single-Case Experimental Designs in Educational Research: A Methodology for Causal Analyses in Teaching and Learning," *Education Psychology Review* 25 (2013): 549–569; Denise A. Soares et al., "Effect Size for Token Economy Use in Contemporary Classroom Settings: A Meta-Analysis of Single-Case Research," *School Psychology Review* 45, no. 4 (2016): 379–399; Stephen F. Ledoux, "Behaviorism at 100," *American Scientist* 100, no. 1 (2012): 60–65; Walter R. Nord, "Beyond the Teaching Machine: The Neglected Area of Operant Conditioning in the Theory and Practice of Management," *Organizational Behavior & Human Performance* 4, no. 4 (1969): 375–401; Laura Vandeweghe et al., "Perceived Effective and Feasible Strategies to Promote Healthy Eating in Young Children: Focus Groups with Parents, Family Child Care Providers and Daycare Assistants," *BMC Public Health* 16 (2016): 1–12.
50. Ledoux, "Behaviorism at 100."
51. Skinner, "A Case History in Scientific Method," 227.
52. Skinner, *The Behavior of Organisms: An Experimental Analysis* (New York: Appleton-Century, 1938).
53. Skinner, "A Case History in Scientific Method."
54. Skinner, *The Behavior of Organisms: An Experimental Analysis*.
55. Skinner, "A Case History in Scientific Method."
56. Alexandra Rutherford, "B. F. Skinner From Laboratory to Life," in *History of Postwar Social Science Seminar Series* (London, 2007).
57. Alexandra Rutherford, "B. F. Skinner's Technology of Behavior in American Life: From Consumer Culture to Counterculture," *Journal of the History of the Behavioral Sciences* 39, no. 1 (Winter 2003): 1–23.
58. Plavnick and Ferreri, "Single-Case Experimental Designs"; T. R. Kratochwill et al., "Single-Case Designs Technical Documentation" (2010). <http://ies.ed.gov/ncee/wwc/Document/229>
59. Plavnick and Ferreri, "Single-Case Experimental Designs"; Hawkins, "The Life and Contributions of Burrhus Frederick Skinner."
60. Skinner, "A Case History in Scientific Method."
61. Ibid., 227.
62. B. F. Skinner, "Superstition in the Pigeon," *Journal of Experimental Psychology* 38 (1948): 168–172.
63. Skinner, "A Case History in Scientific Method," 229.
64. Ibid., 223.
65. Hawkins, "The Life and Contributions of Burrhus Frederick Skinner."
66. Charles B. Ferster and B. F. Skinner, *Schedules of Reinforcement* (New York: Appleton-Century-Crofts, 1957).
67. Skinner, "A Case History in Scientific Method."
68. Ibid.
69. Ibid., 224.
70. Ibid., 224.
71. D. E. Blackman and R. Pellon, "The Contributions of B. F. Skinner to the Interdisciplinary Science of Behavioural Pharmacology," *British Journal of Psychology* 84, no. 1 (1993): 1.
72. Skinner, "A Case History in Scientific Method," 226.
73. Ibid., 224.

74. Ibid., 225.
75. Ibid., 232.
76. Skinner, *The Behavior of Organisms: An Experimental Analysis*.
77. Skinner, "A Case History in Scientific Method," 223.
78. Elisha Babad, Eyal Peer, and Yehonatan Benayoun, "Can Multiple Biases Occur in a Single Situation? Evidence from Media Bias Research," *Journal of Applied Social Psychology* 42, no. 6 (2012): 1486–1504.
79. Ibid., 1501–1502.
80. Diana Baumrind, "Some Thoughts on Ethics of Research: After Reading Milgram's 'Behavioral Study of Obedience'" *American Psychologist* 19, no. 6 (1964): 421–423.
81. Stanley Milgram, *Obedience to Authority: An Experimental View* (New York: Harper & Row, 1974).
82. Milgram, "Some Conditions of Obedience and Disobedience to Authority," *Human Relations* 18, no. 1 (1965): 57–76.
83. Milgram, *Obedience to Authority: An Experimental View*.
84. Diana Baumrind, "Research Using Intentional Deception: Ethical Issues Revisited," *American Psychologist* 40, no. 2 (1985): 65–74.
85. Davide Barrera and Brent Simpson, "Much Ado About Deception: Consequences of Deceiving Research Participants in the Social Sciences," *Sociological Methods & Research* 41, no. 3 (2012): 383–413.
86. Ibid.; D. Geller, "Alternative to Deception: Why, What, and How?" in *The Ethics of Social Research: Surveys and Experiments*, ed. Joan E. Sieber (New York: Springer-Verlag, 1982), 38–55.
87. Barrera and Simpson, "Much Ado About Deception."
88. Karen A. Hegtvædt, "Ethics and Experiments," in *Laboratory Experiments in the Social Sciences*, ed. Murray Webster and Jane Sell (London: Elsevier, 2014): 23–51.
89. Ibid.
90. Baumrind, "Research Using Intentional Deception."
91. Immo Fritzsche and Volker Linneweber, "Nonreactive (Unobtrusive) Methods," in *Handbook of Psychological Measurement—A Multimethod Perspective*, ed. M. Eid and E. Diener (Washington, DC: American Psychological Association, 2004).
92. Ibid.
93. Baumrind, "Research Using Intentional Deception," 166.
94. Philip Zimbardo, "On the Ethics of Intervention in Human Psychological Research: With Special Reference to the Stanford Prison Experiment," *Cognition* 2, no. 2 (1973): 243–256.
95. Philip G. Zimbardo, S. M. Andersen, and I. G. Kabat, "Induced Hearing Deficit Generates Experimental Paranoia," *Science* 212 (June 1981): 1529–1531.
96. G. D. Marshall and P. G. Zimbardo, "Affective Consequences of Inadequately Explained Physiological Arousal," *Journal of Personality and Social Psychology* 37 (1979): 970–988.
97. M. E. Shanab and K. A. Yahya, "A Behavioral Study of Obedience in Children," *Journal of Personality and Social Psychology* 35 (1977): 530–536.
98. Ibid., 534.
99. Ibid., 536.
100. Stanley Milgram, "Liberating Effects of Group Pressure," *Journal of Personality and Social Psychology* 1 (1965): 127–134.
101. Sandra D. Nicks, James H. Korn, and Tina Mainieri, "The Rise and Fall of Deception in Social Psychology and Personality Research, 1921 to 1994," *Ethics & Behavior* 2, no. 2 (1992): 59.
102. Ibid.
103. Ibid.
104. Ibid.
105. J. R. McNamara and K. M. Woods, "Ethical Considerations in Psychological Research: A Comparative Review," *Behavior Therapy* 8 (1977): 703–708.
106. J. Seeman, "Deception in Psychological Research," *American Psychologist* 24 (1969): 1025–1028.
107. Ralph Hertwig and Andreas Ortmann, "Experimental Practices in Economics: A Methodological Challenge for Psychologists?" *Behavioral & Brain Sciences* 24, no. 3 (2001): 383; R. Hertwig and A. Ortmann, "Deception in Experiments: Revisiting the Arguments in Its Defense," *Ethics & Behavior* 18 (2008): 59–92.
108. Hegtvædt, "Ethics and Experiments."
109. James H. Korn, *Illusions of Reality: A History of Deception in Social Psychology* (Albany, NY: University of New York Press, 1997).
110. Baumrind, "Research Using Intentional Deception."
111. Hegtvædt, "Ethics and Experiments."
112. Baumrind, "Research Using Intentional Deception."
113. Ibid.
114. Ibid.

115. Ibid.
116. Barrera and Simpson, "Much Ado About Deception"; S. Bonetti, "Experimental Economics and Deception," *Journal of Economic Psychology* 19, no. 3 (1998): 377–395.
117. Hegtvædt, "Ethics and Experiments."
118. Ibid.
119. Baumrind, "Research Using Intentional Deception."
120. Joan E. Sieber, Rebecca Iannuzzo, and Beverly Rodriguez, "Deception Methods in Psychology: Have They Changed in 23 Years?" *Ethics & Behavior* 5, no. 1 (1995): 67.
121. Baumrind, "Research Using Intentional Deception"; S. A. McLeod, "Psychology Research Ethics," *Simply Psychology* <http://www.simplypsychology.org/Ethics.html>.
122. Baumrind, "Research Using Intentional Deception"; McLeod, "Psychology Research Ethics."
123. Sieber, Iannuzzo, and Rodriguez, "Deception Methods in Psychology," 83.
124. Milgram, *Obedience to Authority: An Experimental View*, 23.
125. Milgram, "Behavioral Study of Obedience," *Journal of Abnormal and Social Psychology* 67, no. 4 (1963): 371–378.
126. Ibid.; Milgram, "Group Pressure and Action Against a Person," *Journal of Abnormal and Social Psychology* 69 (1964): 137–143.
127. Milgram, *Obedience to Authority: An Experimental View*; "Liberating Effects of Group Pressure"; "Some Conditions of Obedience and Disobedience to Authority."
128. Milgram, *Obedience to Authority: An Experimental View*, 27.
129. Ibid.
130. Ibid., 62.
131. Annie Lang, "The Logic of Using Inferential Statistics with Experimental Data from Nonprobability Samples: Inspired by Cooper, Dupagne, Potter, and Sparks," *Journal of Broadcasting & Electronic Media* 40, no. 3 (1996): 422–430; John A. Courtright, "Rationally Thinking About Nonprobability," *Journal of Broadcasting & Electronic Media* 40, no. 3 (1996): 414–421; Michael D. Basil, "The Use of Student Samples in Communication Research," *Journal of Broadcasting & Electronic Media* 40, no. 3 (1996): 431–440.
132. Ibid., 14.
133. Ibid., 16.
134. Ibid., 92.
135. Milgram, "Group Pressure and Action against a Person."
136. Milgram, "Behavioral Study of Obedience," 373.
137. Ibid., 374.
138. Milgram, *Obedience to Authority: An Experimental View*, 203.
139. Ibid., 205.
140. Ibid., 206.
141. Ibid., 22.
142. Ibid., 22.
143. Rebecca S. McEntee, Renita Coleman, and Carolyn Yaschur, "Comparing the Effects of Vivid Writing and Photographs on Moral Judgment in Public Relations," *Journalism and Mass Communication Quarterly* 94, no. 4 (2017): 1011–1030.
144. Philip Zimbardo, *The Lucifer Effect: Understanding How Good People Turn Evil* (New York: Random House, 2007).
145. Ibid., 508.
146. Ibid.
147. Craig Haney, Curtis Banks, and Philip Zimbardo, "Interpersonal Dynamics in a Simulated Prison," *International Journal of Criminology and Penology* 1 (1973): 69–97; Zimbardo, *The Lucifer Effect*.
148. Haney, Banks, and Zimbardo, "Interpersonal Dynamics in a Simulated Prison"; Zimbardo, "On the Ethics of Intervention in Human Psychological Research"; Zimbardo, *The Lucifer Effect*.
149. Zimbardo, "On the Ethics of Intervention in Human Psychological Research."
150. Zimbardo, *The Lucifer Effect*.
151. Zimbardo, *The Lucifer Effect*.
152. Murray Webster and Jane Sell, *Laboratory Experiments in the Social Sciences* (Amsterdam: Elsevier, 2007).
153. Zimbardo, *The Lucifer Effect*, 36.
154. Haney, Banks, and Zimbardo, "Interpersonal Dynamics in a Simulated Prison," 89.
155. Zimbardo, *The Lucifer Effect*, 32.
156. Haney, Banks, and Zimbardo, "Interpersonal Dynamics in a Simulated Prison."
157. Ibid., 21.
158. Ibid., 21.
159. Ibid.
160. Ibid., 24.
161. Zimbardo, *The Lucifer Effect*, 24.

162. Haney, Banks, and Zimbardo, "Interpersonal Dynamics in a Simulated Prison," 73.
163. Ibid., 77.
164. Ibid., 72.
165. Ibid., 69.
166. Ibid., 73.
167. Ibid., 78.
168. Ibid.
169. George A. Miller, "Giving Psychology Away in the '80s," *Psychology Today* 13 (1980): 38ff.
170. Zimbardo, "On the Ethics of Intervention in Human Psychological Research."
171. For an expanded statement of the "so what" question, see the Preface in *The Lucifer Effect*, pages x–xii.

For Review-No Commercial Use(2023)



For Review-No Commercial Use(2023)



THEORY, LITERATURE, AND HYPOTHESES

Full-blown successful interventions never emerge from a brainstorming session. They are always suggested by theory, previous research, or extensive clinical experience.¹

—R. Barker Bausell

For Review-No Commercial Use(2023)

LEARNING OBJECTIVES

- Explain the role of theory and literature in experiments.
- Prepare a literature review that connects previous work to your own and makes a theoretical contribution.
- Describe the role of hypotheses and research questions in experiments.
- Summarize when to predict differences or direction in hypotheses.
- Formulate original hypotheses using the diagram given.

To Bausell's assertion, we should add that experiments are not **exploratory** in the sense that they are used when problems are in the preliminary stage.² Rather, experiments presume a fair amount of well-developed theory and evidence from previous research. Thus, it is important to do the homework before starting any experiment, which involves a thorough search of the literature, including theory. This



iStock.com/haramee2554

chapter builds upon the previous one by examining the role of what researchers have discovered before us, or Isaac Newton's concept of "standing on the shoulders of giants."³ While the information in this chapter also applies to other research methods, it is especially true that experiments are designed to create new knowledge by building upon what has been discovered previously. It also addresses two of the seven "attributes of a well-executed experiment."⁴

That includes theory **explication**, or explaining the theory being tested, and clear identification of research questions and hypotheses. It looks at the literature review section of the paper from the perspective of how to explain the theory being tested, relate it to the variables used in the experiment, and link all the hypotheses to that theory. Hypotheses are particularly powerful analytic tools in experiments, so knowing how to write a literature review and the hypotheses that arise from it is especially important for this method.

As covered in chapter 1, the first step in designing an experiment is to formulate an idea of what causal relationships are to be tested and clearly articulate that in the statement of purpose. But these relationships should not be based merely on hunches or curiosity. Preferably, the idea of what to test should grow out of an understanding of the domain and the literature and theories within it. Measuring causality requires thinking theoretically; otherwise, experiments would be a haphazard process of trial and error.⁵ (For more about theory, see More About box 3.1.)

One common pathway by which social scientific knowledge can develop is by starting with exploratory work using methods such as observations, interviews, and ethnography, among others. These types of qualitative methods are explicitly designed to be descriptive and generate theories.¹⁴ They reveal **latent** meaning—that is, not readily apparent on the surface—or help us understand the nature of phenomena and predict the conditions under which they occur.¹⁵ But because of their inherent qualities (small samples, not randomly drawn, lack of control, etc.), they are not well suited for testing or refining the theories they generate. That is one of the main purposes of experiments.¹⁶ After theory is generated by qualitative or other means, research then can progress to methods that test them. Surveys are good for eliciting opinions, attitudes, and self-reports about behavior. Content analyses can uncover the characteristics of various types of messages, text, or visuals. While these methods produce research that adds new knowledge and expands theory, they also suggest causes and provide correlational evidence that lays the groundwork for experimental work. Experimental studies are then able to determine if the phenomenon observed in a natural setting can be reproduced in a controlled environment,

MORE ABOUT . . . BOX 3.1

Theory

Theory is a generalized explanation for some particular phenomenon, or “an organized body of concepts and principles.”⁶ This is an academic way of saying that theory is about things that actually occur in the real world. To invoke Kurt Lewin, the father of gatekeeping, “There is nothing more practical than a good theory.”⁷ Although theory can seem mysterious and hard to comprehend, it really is nothing more than a tool for understanding what happens in life. Theories evolve when hypotheses are continually supported.⁸ A good theory should do more than describe; it should also explain and allow predictions about the phenomenon. For example, one aspect of framing theory describes how news stories are constructed as discrete episodes—a story about a crime, for example—or more generally as themes that add context—a story about crimes, not just one in particular. It explains how people perceive these types of frames and also allows us to predict how people will react to them—for example, by holding individuals responsible when events are framed episodically but placing blame on social factors when events are framed thematically.⁹ Experiments build theory by testing assumptions and expanding upon our existing understanding.

However, there has been some discussion of “exploratory experimentation” in biology and related disciplines.¹⁰ What these researchers mean by an exploratory experiment is not an atheoretical, aimless data-gathering project. Rather, an exploratory experiment is one that is *directed* or *informed* by theory but does not necessarily test it via formal hypotheses.¹¹ These studies have a theoretical background, not a lack of theoretical framework altogether.¹² Theory is still “crucial to the success of exploratory experiments.”¹³ It is never appropriate to justify a lack of theory, low power, subjects not randomly assigned, uncontrolled variables, or other weaknesses by labeling the experiment as “exploratory.” In these cases, the researchers should consider it to be a pilot study, take what they learned, and mount another experiment that improves upon the weaknesses rather than attempt to publish it as an exploratory work.

where all the many variables that could confound results or be responsible for plausible alternative explanations in the real world can be ruled out. Having arrived at the place where an experiment is the next logical method, the first step is to conduct a thorough search of the literature and particular theory.

All this is not to say that researchers can never come up with ideas for experiments by other means, such as by observing and becoming curious about why a particular phenomenon happens, as James Lind did for his study of scurvy. Ideas generated this way can certainly be tested with experiments. But there also should be some reason for testing them based on theory or evidence. When pursuing research questions generated from

observations, it is important to find empirical evidence and theories that can help guide and refine them. This purpose can be served by correlational studies that link two variables, as can experiments that use a different population or different dependent variable (DV). For example, in a study to see if journalists' moral judgment could be improved by showing them photographs,¹⁷ no work had been done using moral judgment as the DV and photographs as the independent variable (IV). However, photographs had been shown to improve other cognitive processes such as elaboration, empathy, and involvement.¹⁸ All of these DVs are related to moral judgment, as shown by other studies.¹⁹ These theoretical and empirical rationales provided links to the independent and outcome variables, albeit indirectly. This idea came about as a result of much speculation in essays and other nonempirical works about how journalists' ethical reasoning improved when photographs were involved.²⁰

THE LITERATURE REVIEW

The explanation of the theory being used and how it relates to the variables of the study is achieved in the literature review section of the paper. A well-done literature review provides a strong foundation for the entire experiment, yet Bausell says the “single rule most often violated by beginning researchers” is not knowing the relevant literature thoroughly enough.²¹ Seasoned researchers may not feel the need to write a formal literature review in advance of running an experiment because they keep abreast of the latest findings and also work with theories they know well. But for others, before forming any testable hypothesis, a thorough search of pertinent literature and familiarization with relevant theories should be undertaken. The first step in this process is to perform a search of the literature, and the second is writing it up in the literature review section of the paper, which is the subject of this chapter.

Reading others' research can be a great source of ideas for experiments. First, reading existing studies will show whether the idea has been done already, so the study is not just a simple replication.ⁱ It will also reveal gaps in what has been done and what still needs to be done to flesh out a theory. The statements of “limitations” and “future studies” at the end of every research article can provide even more ideas. This includes ideas for studies that need replicating and for extending and improving upon a theory. Reading literature also provides methodological help, including information about how others have measured the variables, how many subjects might be needed, and ideas for novel procedures.

ⁱIn some disciplines, replications that do not extend theory or empirical knowledge in some relevant way are frowned upon.

The screenshot displays a database search interface. At the top, there are navigation tabs: 'New Search', 'Publications', 'Subject Terms', 'Cited References', and 'More'. On the right, there are links for 'Sign In', 'Folder', 'Preferences', and 'Language'. The main search area shows 'Searching: Academic Search Complete | Choose Databases'. The search query is 'experiments AND social science'. Below the query, there are options to 'Select a Field (option...)' and buttons for 'Search' and 'Clear'. There are also 'AND' and 'OR' operators and a 'Limit To' section. The search results are displayed in a list format, showing the first 10 of 9,003 results. The first result is '1. Undertaking Experiments in Social Sciences: Sequential, Multiple Time Series Designs for Consideration.' by Phan, Huy; Ngu, Bing. The second result is '2. Field Experiments Across the Social Sciences.' by Baldassarri, Delia; Abascal, Maria. The interface includes a 'Refine Results' sidebar on the left with options like 'Current Search', 'Boolean/Phrase', 'Limit To', and 'Publication Date'.

For Review-No Commercial Use(2023)

This book presumes that readers are familiar with how to conduct a literature search and write a review, perhaps even having conducted and written one or more already. Because an experiment is so much work on the front end of designing it, and it cannot be changed once the data have been collected, it is especially important that the literature search for an experiment be performed thoroughly. I recommend actually writing a draft of the literature review before proceeding to collect data no matter how familiar you are with the topic. I have found that the act of writing slows me down and gives me time to think about the literature in a way I would not if I had just been reading it. It also forces me to synthesize the many different findings, which helps me better analyze the meaning and find gaps that may need filling. Making a chart or table can help show things more clearly, and sometimes these figures end up in the published study—for example, table 3.1.

With experiments, it is necessary to isolate or control all things that could possibly affect the outcome or DV except for the one thing being manipulated, so it is especially crucial to know what all those things are. They can then be controlled, or used as mediators, moderators, or covariates. If a researcher does not know what they are and fails to measure or control them, the validity of the findings for the entire experiment is put in jeopardy.

There are many good books and articles on how to perform a literature search and write a literature review, so those will not be described in detail here. Instead, next is advice for avoiding some of the most common problems with literature reviews in

TABLE 3.1 INDICATORS OF IMAGINATIVE CAPACITIES (LIAO ET AL., 2016)		
Indicator	Definition	References
Novelty	The ability to generate unique ideas	Beaney (2005) and Vygotsky (2014/1967)
Productivity	The ability to produce thoughts using the extensive application of free association	Folkmann (2010) and Gaut (2005)
Concentration	The ability to formulate thoughts through focus and immersion	Csikszentmihalyi (1996) and Folkmann (2010)
Sensibility	The ability to evoke feelings during the creative process	Ricoeur (1978) and Scheffler (1986)
Intuition	The ability to generate immediate associations with a goal	Reichling (1990) and Townsend (2003)
Effectiveness	The ability to generate relevant and profound thoughts to attain a goal	Gilbert and Reiner (2000) and Shin (1994)
Dialectics	The ability to seek improvement by logically analyzing possibilities and alternatives	Cartwright and Noone (2006) and Reiner and Gilbert (2000)
Exploration	The ability to inquire about the unknown	Coiello (2007) and Thomas (2004)
Crystallization	The ability to visualize abstract concepts by using concrete examples	De Vries (1988) and Vygotsky (2004/1967)
Transformation	The ability to perform tasks by applying information acquired across multiple fields of knowledge	Kunzendorf (1982) and Liu and Noppe-Brandon (2009)

From: Liao, Kai-Hung, Chi-Chen Chang, Chao-Tung Liang, and Chaoyun Liang. 2016. "In Search of the Journalistic Imagination." *Thinking Skills and Creativity* 19: 9–20.

experimental designs. These tips are not exclusive to this method, but the problems they address occur often enough in experiments that they are worth mentioning.

Tips on Writing the Literature Review for Experiments

It Is Not a Book Report

It is important to analyze and synthesize the findings of the various studies found in the literature search. Too often, literature reviews read like book reports. They are summaries of the findings of previous studies, but there is no analysis of those findings that shows the connections among or between them and the study about to be conducted. Do not assume that readers will be familiar with the topic; therefore, how the studies reviewed connect to each other and to the current study will not be

as obvious to them as to you. Spell it out. Furthermore, experiments test hypotheses, which are proposed based on previous findings; therefore, an analysis of those findings is especially important for this method.

Rather than stringing together summaries of the abstracts (Smith and Jones found . . . ; White, Brown, Green, and Black showed . . . ; Tinker and Taylor studied . . .), a good literature review is a critical analysis of the studies, synthesizing the findings, and coming up with new ideas of your own. This approach also avoids what can look like shameless name-dropping or an attempt to bump up an author's citations or a journal's **impact factor**—a measure of how important it is.

As an example of analyzing and synthesizing, I relate a story of a colleague and I who were doing an agenda-setting study to discover whether issues in the news or attributes of candidates had more powerful effects on voters. We conducted a literature search and noticed that the correlations between the media's agenda and the public's agenda (the basic premise of agenda-setting theory) were consistently higher for attributes than for issues. That was the foundation for the resulting hypotheses, which predicted that attribute agenda setting would be more powerful than issue agenda setting.²² This was not something we found already written about but discovered by thinking critically about the theory and comparing the findings of all the studies.



Stock.com/Wallpaperpress

Do Not Be Wikipedia

Too many literature reviews read like Wikipedia entries. Instead of reviewing everything ever written about a topic, a good review covers only the literature related to the study. For example, if the experiment is on second-level agenda setting, it might be important to briefly mention what first-level agenda setting is and say how it differs from second level, but it is not necessary to explain it in detail, summarize every study on it, and describe all the mediators, moderators, or **contingent conditions**—that is, when it works and does not. Instead, mention it and say how it differs from what this experiment will test, and move on. In other words, only focus on B if the study is about B, even if there is also an A. Do be sure to cover all the *relevant* literature; some authors go too far in the opposite direction and leave out important work. Key is finding balance—cover just the right amount and do not leave out any of the truly **seminal studies**, or the classics.

This also applies to the theories written about in the literature review. There are many theories that overlap and can be used to inform an experiment, but a study should

STUDY SPOTLIGHT 3.2

Shoemaker, P., J. Tankard, and D. Lasorsa. 2003. *How to Build Social Science Theories*. Thousand Oaks, CA: Sage.

Novice researchers frequently have a difficult time understanding not only what theory is but especially how to build it. It is not uncommon for journal reviewers to ask authors: “What is your theoretical contribution?” “What theoretical statements does this study offer?” and “What does this study do to build theory?” Finding ways to answer these concerns keeps many a budding social scientist up at night. When this book on how to build theory in the social sciences came out, it spoke directly to my inner experimentalist in its logical, step-by-step approach. It delivered on its promise to teach “the challenging activity of theory building with the minimum amount of difficulty.”²⁵ This book begins by assuring readers that they are not alone and that there is no shame in admitting a lack of understanding of how to build theory and provide theoretical contributions.

The chapters build logically, beginning by describing hypothesis writing. It then explains theoretical concepts, which it calls “the building blocks of theory.”²⁶ It illustrates ways to combine concepts into theoretical statements, which broadly describe relationships among variables. It shows how to synthesize disparate parts and make linkages, both theoretical and operational. Then it shows how to describe and explain the relationships among concepts in ways that offer explanatory power. Chapters deal with how to relate two variables, three, four, and more, and how to build models in advance of a full-blown theory.

The book offers examples that social science scholars can relate to. The development of cultivation theory by George Gerbner and colleagues is one, showing how it began as a way of explaining the effects of television on viewers.²⁷ One hypothesis proposed that watching television shapes viewers’ beliefs, ideologies, and worldviews. As studies were conducted and the hypothesis supported, it grew to include other concepts with labels such as “resonance” and “mainstreaming,” which specified the conditions necessary for the effect to occur. More research resulted in more revisions and additions to the theory, including first- and second-order effects that delineated perceptions based in reality and beliefs.

How to Build Social Science Theories stresses the importance of creativity to building theory, an idea that is also important in this book. It offers concrete, practical ways to make contributions to theory. How to develop theory never really “clicked” for me until I read it.

typically use only one or two. For example, there are at least ten theories that can be used to explain aggression from watching television.²³ In studies using other methods, reviewing many similar theories may be more appropriate than in an experiment designed to test a specific theory. Find the one theory that is best suited to what the experiment is trying to do, and write a solid rationale for why it is used and not others rather than reviewing every possible theory. Sometimes, two theories will lead to opposite predictions, and so those can be juxtaposed in a single study.²⁴ In this process, a compelling link should be made between independent, dependent, and **intervening variables**, or those that come between the treatment and the outcome.

Make a Theoretical Contribution

Far too many studies of all kinds will summarize a theory and then go on to do whatever the authors intend to do but never circle back to the theory at the end. They often fail to explicitly say what the study intends to do to contribute to that theory or theories. With experiments, this is an especially egregious oversight given that one of the main reasons for doing an experiment is to develop, test, refine, or otherwise contribute to theory. It is never enough to lay out a theory as the foundation and then abandon it. It should be explicitly stated what this experiment will do for the theory or theories, and that should be expanded on in the discussion and conclusion of the paper. (For a synopsis of a book on how to build theory, see Study Spotlight 3.2.)

Toward this end, it is important to link all the variables—dependent, independent, causal mechanisms, and individual difference variables—clearly explicating the relationships among them.²⁸ Start by stating how the IV will affect the DV, and then explain through what theoretical processes this will occur. If it helps, make a figure that models this and use it to visualize the process (for an example, see table 3.1). As an example, the study on photographs' effects on moral judgment²⁹ mentioned earlier was interested in establishing a causal relationship between journalists who were exposed to photographs and their levels of moral judgment compared to journalists who were not exposed to photographs. The study theorized this would occur because photographs had been shown to increase empathy, which intensifies involvement and also encourages central route processing, all of which were found to increase moral judgment. The theoretical contribution was not only that photographs had the ability to improve moral judgment—an intervention that had not been tested previously—but also the demonstration of the theoretical explanation for *why* photographs worked.

Connect the Dots

Finally, another important feature of a good literature review is to relate other studies to your own, what I call connecting the dots. After reviewing every batch of studies on a common topic, explicitly say how they are related to the study to be conducted. Will it be measured the same way? Do they contain gaps that this one will fill? What connections in these studies will this experiment build upon? Always relate the findings of other studies to your study, locating it within the context of the larger body of work. A reader should never have to go more than a paragraph or two to find out why specific studies are being reviewed and how they are connected to the current one. For example, in the study mentioned earlier that examined the correlations from various agenda-setting studies, the paper explained how they were relevant to the current study by saying, “All these second level correlations are similar to the highest

HOW TO DO IT 3.3

Examples of Connector Sentences

In a literature review, it is important to relate studies to each other and to your own, locating your work in the context of others. I offer here some examples of how various researchers have made this explicit in their experiments.

- After reviewing the three sequences in the Hierarchy of Effects Model, this study said, “We are not particularly concerned with the sequence but rather with the strength of the first two components of these models—knowledge and attitudes—on the behavioral outcome, voting. In other words, we seek to determine not which came first, but which has a stronger effect: cognition about issues (first level) or feelings about attributes (second level).”³¹
- After reviewing celebrities’ Twitter use and source credibility, the authors say, “Acknowledging the aforementioned novelty and importance of celebrity-generated messages embedded in social media, this research tested the effects of celebrities’ Twitter-based electronic word-of-mouth on consumers’ perceptions of source credibility, intention to spread electronic word-of-mouth, brand-related outcomes, online bridging social capital, and social identification with celebrities.”³²
- Before reviewing attribution styles, this study said, “This study delves into the role of consumers’ different attribution style, specifically how different attribution styles lead to the differential impacts in revealing the role of the negative presentation of information.”³³
- In a study of math learning, the authors reviewed the literature and summarized the findings, and then tied it to the question this study was designed to answer this way: “Simply put, as math anxiety increases, math achievement declines. A possible inherent relationship between anxiety and achievement poses an obvious question, however: Is a poor performance on a math assessment/problem due to math anxiety or due to lack of mastery of the content?”³⁴
- After reviewing studies on different methods that were effective in helping students learn English, this study summed it up this way: “An examination of the salient characteristics and benefits of a technology-enriched curriculum for English Language Learners underscores the pivotal role Computer Assisted Language Learning can play in second language teaching and learning.”³⁵
- Before summarizing literature, the authors of a study testing businesses that drop out of overcrowded markets note the similarities among the studies and point out the gap their work fills: “Although wars of attrition have an important place in the game theoretic literature, there are surprisingly few experimental studies directly relating to them.”³⁶

correlations at the first level; none reaches the lowest levels or even the mean level of .53 in meta-analysis of the first level effects.”³⁰ That is an example of a sentence that connects all the studies just reviewed to the current one.

These connections can never be too obvious. Good connecting sentences include: “This is relevant to our study because . . .” “We use this finding in our study by incorporating it as a mediator . . .” “These studies raise questions about . . . , which is the purpose of this research,” and similar phrases. Many good examples of connector sentences exist; some are presented in How To Do It box 3.3. As you read other studies, look for them and adapt to your own work.

Conceptual Definitions vs. Operationalizations

The literature review is also where all the important concepts and variables in a study are conceptually defined. A **conceptual definition** is an abstract, theoretical description of something using general qualitative terms. Students often confuse conceptual definitions with **operational definitions**, or operationalizations, which take a concept and define it in the specific, concrete ways it is measured. A **concept** is a general idea about something that has many specific characteristics. For example, credibility was conceptually defined as “the public perception of news quality”³⁷ in one study, and then it was operationalized with thirteen specific characteristics that included how fair, complete, accurate, believable, credible, informative, interesting, likeable, in-depth, important, and well written a news story was, plus how trustworthy the sources and information were. These are two different kinds of definitions, which appear in two separate sections of a paper. It is not adequate to provide a list of terms to be used to measure the concept as the conceptual definition. The conceptual definition should instead describe and explain the concept based on theory. Another example is the conceptual definition of moral judgment as the reasons “that people use to decide that a course of action is ethically right or wrong.”³⁸ The operationalization given in the methods section is a list of twelve statements about things that were important to the subject when deciding what to do about a particular dilemma (for examples, see the footnoteⁱⁱ).

ⁱⁱThe twelve statements vary according to each dilemma, but an example is: (1) A chance like this photo comes only a few times in a career; (2) The kids will grow up having this horrible, graphic reminder of what happened; (3) This is a family newspaper, children might see this photo at the breakfast table; (4) You'll probably get a lot of angry calls and people will cancel their subscriptions; (5) How these people and their families will feel when they see this; (6) Your competition is working on a similar story. If you don't run the photo, your competition will just run something similar; (7) Publishing this photo would help your paper's reputation for investigative reporting; (8) Whether it is our duty as journalists to show all the news, regardless of the circumstances; (9) Whether the public has a right to know all the facts about drug use and its effects on people, especially children; (10) What would best serve society; (11) Photos that are painful to some have to be shown so others will benefit; (12) If I don't run this photo, I may prevent these children from being taken from their parents, but the conditions leading to situations like theirs will persist.

The credibility study cited previously also provides a good example of a concept that has been defined differently by various researchers, including as “attitude toward a source of communication,” “global evaluation of the believability of the message source,” a combination of trustworthiness, expertise, and goodwill, and as the combined public perceptions of media’s approach to “quality, profit making, privacy, community well-being and trustworthiness.”³⁹ It is not uncommon to find a lack of consensus about the conceptual definition of some concept. When that is the case, it is important to acknowledge that in the literature section, review the different conceptual definitions, and then defend your choice of definition and explain why it is best suited for your study. To repeat: Every variable used in a study must appear in three places:

- conceptually defined in the literature review,
- incorporated in the hypothesis or research questions, and
- operationalized in the methods section.

Trace each variable throughout the paper and be sure that each one appears in all three places. There should be no orphan concepts or variables.

For Review-No Commercial Use(2023)

A good way to make sure that all bases are covered is to think of the literature review in terms of an outline that lists all the DVs, IVs, intervening variables (mediators, moderators, causal mechanisms), and covariates, if any. Use subheads that mirror the outline to help clarify it for readers and also to help organize the writing. This ensures that everything is defined, the empirical evidence is covered, and all the discussion about one variable is in the same place. Experienced readers expect to find all the information about a particular concept together and can become annoyed when something pops up in different places throughout the paper.

Literature Reviews With Multiple Experiments

Many researchers report more than one experiment in a single article for reasons that range from replicating a finding to ruling out plausible alternative explanations. Sometimes this is planned beforehand, and other times the need arises only after the first study results are known. This phenomenon will be explored in chapter 5 on validity. When multiple experiments are reported in one paper, there may be one section that covers all the literature for both studies, or there may be two literature reviews, the first one the longest and the second that includes only new literature that pertains just to the second study. Usually, these will be labeled “Study 1” and “Study 2,” or more if there are more

experiments. The general pattern is to report the literature for the first study, followed by the methods for it under a subhead such as “Study 1,” then the results and discussion of it. The second study will follow under a subhead “Study 2” or “Experiment 2,” starting with a description of it and a small literature review if necessary to explicate any new variables. For example, if new IVs, mediators, or moderators were introduced in order to rule out confounds or plausible alternative explanations that may have arisen in the first study, they obviously were not covered in the first literature review. It is not necessary to repeat all the previous information from the main literature review, only to add the new concepts and put them into context with the previous literature and theory. A good illustration comes from a study of cognitive processing in agenda setting.⁴⁰ The second experiment was conceived of after the first one showed the effects the authors expected on some subjects but not others. They speculate this could have been affected by the content of the stories they used, and write: “We decided to conduct a follow-up study to test the assumption that people who process information more centrally than others could be influenced by content aspects of the articles such as journalistic evaluations of an issue’s importance.”⁴¹ They follow with the subhead “Study 2” and begin with a one-page literature review just of studies that deal with content of news stories.

Finally, a few miscellaneous but still important points for the literature review:

For Review-No Commercial Use(2023)

- *Include current sources.* In addition to the seminal or classic studies, be sure to include up-to-date works. This is especially important if a study was started some time ago. Go back and see what has been done recently and include it.
- *Use different search methods.* Searches using key words in databases do not always uncover all the important literature, even when using multiple databases. To avoid missing important studies, go through the last two or three years of journals that are likely to publish articles on the topic to see if anything has been missed. Include the journal you plan to submit to in this process.
- *Include conflicting findings.* Very few programs of research always produce consistent findings. Look for studies that *did not* work as expected. Not only is it important to be honest, it is important to know what might cause your experiment to not work as planned so you can control for it as much as possible. Review the studies with findings that do not support your predictions and explain why they might have come out this way.
- *Use quotes sparingly.* Quoting others at great length or too often raises red flags for readers because it can appear as if you did not understand what the

author was trying to say. Paraphrasing someone else's writing also affords the opportunity to write it in a clearer way. Make it a habit not to use direct quotes unless there is absolutely no better, clearer, more concise way to say it. There are very few sentences in academic papers that meet this criterion. When direct quotes are used, cite them appropriately with page numbers.

- *Be careful not to plagiarize.* See preceding point. Much plagiarism results from careless practices rather than bad intentions—for example, taking notes by copying and pasting from the original article and then forgetting whether these were already rewritten in your own words. Repeating your own writing from a different article is also a form of plagiarism. Develop a system to know what is paraphrased and what is not, for example, by highlighting, using quotes, or putting them in a different type font.
- *Always read the paper.* It can be time consuming to trace back and read the original papers cited in others' literature reviews, but relying on another author's interpretation of a paper rather than reading it yourself can be perilous. Scholars have been known to uncritically report the same summary of a study over and over, perhaps even wrongly.⁴² I find this to be the case particularly when books

For Review-No Commercial Use(2023)

HYPOTHESES AND RESEARCH QUESTIONS

The literature review is also where the hypotheses and research questions are typically incorporated. One approach is to list all hypotheses at the end of the literature review. This style involves a block of hypotheses before the methods section, like this:

“Hypothesis 1 (H1): Individual donors will donate less when they know that a nonprofit receives some amount of government funding. (Categorical crowding-out hypothesis)

Hypothesis 2 (H2): Individual donors will donate less if a nonprofit receives a greater share of its funding from government. (Continuous crowding-out hypothesis)

Hypothesis 3 (H3): Knowing that government funding comes from a competitive merit-based program will increase individuals' willingness to donate. (Crowding-in hypothesis)”⁴³

For the purpose of illustration, in these hypotheses I have slightly edited the second one to keep the wording parallel; the original paper uses “give” in H2, while the other hypotheses use

the term “donate.” As “give” and “donate” could conceivably be defined differently, it is important not to use synonyms in writing hypotheses. Instead, use parallel wording for hypotheses whenever possible.

To keep readers from having to refer many pages back in the literature review in order to recall the evidence leading up to each hypothesis, some journals use a short summary of the literature that led to each prediction just prior to or after the hypothesis. As it can be inefficient to have the full description of studies in the literature review section and then to repeat a summary of it at the end of the literature review with the hypotheses, another style embeds hypotheses into the literature review at the point where the topic of each hypothesis is covered. So, for example, literature on topic A is followed by a hypothesis related to A; literature on topic B is followed by a hypothesis related to B. That approach allows the empirical evidence to be fresh in readers’ minds. For example, in a study on the climate of opinion judgments comparing the relative weight of explicit cues to implicit cues,⁴⁴ the authors first reviewed the literature on explicit cues and gave this hypothesis:

For Review-No Commercial Use(2023)

“H1: Surveys, as explicit media cues, influence recipients’ judgments of the climate of opinion in the direction of the survey results.”⁴⁵

They next reviewed implicit cues and predicted:

“H2: If arguments in a media report support (oppose) a certain opinion, the perceived public agreement to this opinion increases (decreases).”⁴⁶

They then reviewed literature on persuasive effects of both kinds of cues in media and gave this hypothesis:

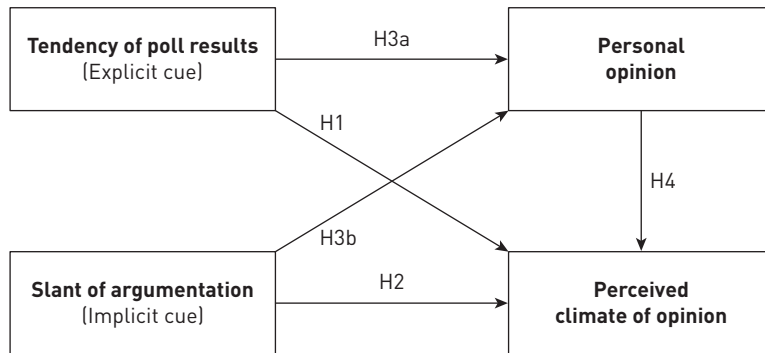
“H3a: If survey information on an unknown issue is presented, recipients tend to follow the majority opinion.”⁴⁷

There were several more hypotheses, with it all visually summarized in a model (see figure 3.1).

It can be a challenge not to introduce a concept in a hypothesis before having explicated it in the literature review. So, for example, do not introduce hypotheses that include the



FIGURE 3.1 A MODEL SHOWING THE IVS, DVS, AND PROCESS PROPOSED IN HYPOTHESES



Source: Zerback, Thomas, Thomas Koch, and Benjamin Kramer. 2015. "Thinking of Others: Effects of Implicit and Explicit Media Cues on Climate of Opinion Perceptions." *Journalism and Mass Communication Quarterly* 92 (2): 421–443.

current or future state of something until the empirical evidence on that has been covered. Organizing the evidence in a way that follows naturally and does not assume reader knowledge of concepts before they are explicated is tricky. This makes it tempting to plop all the hypotheses in at the end, but then readers may have trouble recalling what was said about the concepts in H1 by the time they get to it. Instead, breaking hypotheses down into smaller statements, one IV at a time, in the order that the process occurs, is key.

The hypotheses in the previous examples are formally phrased and given the designation “H” plus a number (H1, H2, H3, etc.). The lower-case letters (H5a, H5b, H5c) indicate a subset of related hypotheses. But it is also quite common to see hypotheses phrased in a narrative style. For example, in a study of whether authoritarian governments are willing to include citizen preferences in the policies they enact, scholars in China posed their hypotheses this way:

Thus, based on previous research as well as statements by the CCP, we have reason to believe that when state–society relations are harmonious, leaders may be equally receptive to opinions expressed on the Internet and to those expressed through formal channels. However, when leaders believe that antagonism exists between the state and citizens, we expect that they will be less receptive to both formal and Internet channels.⁴⁸

Hypotheses should be written in future tense (e.g., “will have,” “may be”) rather than present tense (e.g., “is”) because they have yet to be tested; the present tense can sound like it is already a fact.

How many hypotheses are needed? Students often ask this question, but I am unable to give anything but the Goldilocks answer: Not too many, but not too few—just the right amount. One hypothesis may indicate a too-small study or the need to break a too-large hypothesis down into smaller bites. Hypotheses that run into the double digits may indicate a study that attempts to do too much and should be split into two publications. Let theory, logic, and reader attention decide. Have colleagues read the literature review to see if they can follow it clearly without getting lost or bored.

More important than how many hypotheses there are is how clearly they are written. If a process is specified, the hypotheses should build logically upon each other. Sometimes, experiments explore causal mechanisms that involve multiple steps in complex theoretical processes. When this is the case, it makes sense to break things down into a series of specific hypotheses that build upon each other in a cohesive model. For example, in the study of how photographs elevate moral judgment, the process was broken down into three hypotheses:

H1: Participants who see photographs will have significantly higher levels of elaboration about stakeholders than those who do not.

H2: Moral elaboration about stakeholders will be significantly associated with higher ethical reasoning.

H3: Ethical reasoning will be significantly higher for participants who see photographs than for those who do not.”⁴⁹

The third hypothesis was what the study was most interested in, but it occurred last in the causal chain of events.ⁱⁱⁱ

Mutz offered a similar set of hypotheses, in narrative fashion, in a study of close-up camera perspectives and uncivil TV discourse (they are broken into separate paragraphs for easier reading here):

“The first hypothesis is that close-up camera perspectives and incivility will both increase levels of emotional arousal.

A second hypothesis following from the first is that heightened arousal will increase levels of recall.”⁵⁰

ⁱⁱⁱThis study also predicted an interaction of photographs and involvement, but I leave the discussion of main effects and interactions for later.

In this example, the first prediction involved close-up camera perspective and incivility leading to emotional arousal. The second prediction takes arousal from the first hypothesis and connects it to recall. The order is not chosen randomly; the researcher proposed them in this order because it was posited that they occurred in this order. With more involved processes, these pathways can be diagrammed out in figures to make it easier to visualize.

Hypothesis Basics

Null vs. Alternative

Before explaining how to write a hypothesis, this is a good place to review the concepts of null and **alternative hypotheses**, where a null hypothesis says there is no difference or direction, and the alternative hypothesis says there is. These were invented in order to reduce experimenter bias, or the tendency to find evidence that supports our ideas. A long time ago, scientists came up with the idea of basically arguing against your own position—that is, researchers should not try to show that a treatment causes effects but to *disprove* that it does not. In other words, to try to show that there is no relationship between two things, or no difference between two groups. The null hypothesis is written H_0 and the alternative H_1 or H_a . By rejecting the null, we say we have reason to think that there is a relationship or a difference, thus we can say there is a likelihood that our hypothesis is correct. This is technically the alternative hypothesis. In other words, we start from the position that the null (no difference, or no relationship) is true until evidence shows otherwise. If the null hypothesis is rejected, we say we have found support for our hypothesis. We never use the term *prove* because that cannot technically be done. Instead, the proper language to use is to *support* our hypothesis or *reject* the null; to reject the hypothesis or have it “*fail to be rejected*”; or to say that the hypothesis was “*not disconfirmed*.”⁵¹ This is rarely stated explicitly in research papers but is the thought process underlying them all.

A hypothesis is a prediction of how independent and dependent variables are related.⁵² They can be thought of as “if . . . then . . .” statements—for example, “If journalists see photographs, then they will use higher levels of moral judgment.” Hypotheses should be stated in terms of concepts rather than operationalizations.⁵³ For example, in business, there is a concept called managerial trustworthiness, abbreviated MTW, and conceptually defined as “the trustworthiness attributed to supervisors.”⁵⁴ It is measured with nine items including capable, competent, concerned, and a strong sense of justice, among others. The hypotheses refer to managerial trustworthiness, not competence, justice, etc. For example:

“H1: Emphasis on internal management that relates to setting challenging but feasible goals has a positive effect on an individual’s perception of MTW.”⁵⁵

Difference vs. Direction

Hypotheses come in two flavors: those that predict a causal direction—for example, which of two variables will have larger effects than the other—and those that merely predict a difference but do not propose which variable will have a greater impact than the other. These are also called *nondirectional hypotheses*. An example of a hypothesis that predicts a difference is:

H: Journalists who see photographs will use significantly *different* levels of moral judgment than journalists who do not see photographs.

It predicts a difference between seeing and not seeing photographs, but it does not specify whether the outcome will be higher or lower. By this account, seeing photographs could result in moral judgment that is either better or worse. By contrast, this version of the hypothesis predicts a direction:

H: Journalists who see photographs will use significantly *higher* levels of moral judgment than journalists who do not see photographs.

The second hypothesis goes further than the first to say specifically what the difference between the two conditions—seeing photographs or not—will be.

Another example of a directional hypothesis is from the previous example of a study on managerial trustworthiness:

“H1: Emphasis on internal management that relates to setting challenging but feasible goals will have a positive effect on an individual’s perception of MTW.”⁵⁶

This hypothesis predicts a causal direction with the use of the word *positive*; the researchers expect *higher* levels of MTW in the goal-setting condition. A nondirectional way to state this hypothesis would be to say that goal setting would have an effect on MTW that was different from the other conditions, but not to specify whether that effect would be positive or negative.

Whether to make a hypothesis that predicts a direction or a difference depends on the theory and empirical evidence available; if there is enough to suggest a causal direction, then it is appropriate to make that prediction. If there is not enough evidence, or theory does not suggest a direction, then it is more appropriate to predict a difference but not specify a direction. For example, the first time the effects of photographs on moral

judgment were tested, there were theoretical reasons to believe that photographs could either improve or worsen moral judgment. In that case, it is appropriate to propose that seeing photographs will cause subjects to behave differently without predicting whether their moral judgment would be better or worse than those who did not see photographs. After a study finds effects in a certain direction, from then on it is appropriate to use directional hypotheses.

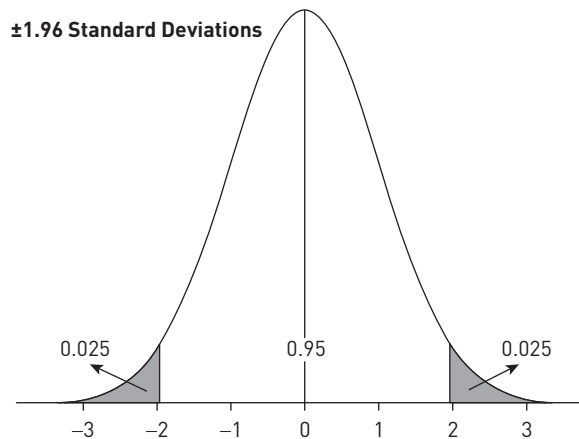
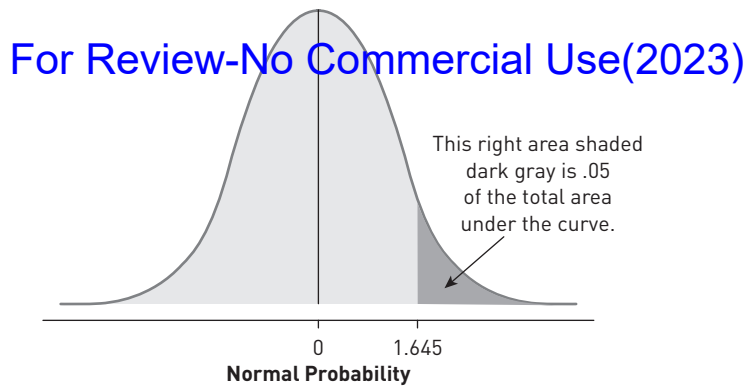
If theory and prior evidence is in such short supply as to prevent a researcher from making a prediction at all, then a **research question** should be used. A research question represents a more preliminary state of affairs concerning theory and evidence.^{iv} So continuing the previous example, if there was even less theory and evidence to make the prediction that photographs would have some effect on moral judgment, it is appropriate to ask, “Does seeing photographs affect journalists’ moral judgment?” As a rule of thumb, it is always best to make the most specific prediction possible with the available evidence and theory, as that provides us with more powerful analytical tools.⁵⁷

Whether a researcher makes a prediction of direction or not is important in how hypotheses are tested statistically. What follows may not make sense to readers who have not taken an introductory statistics course; for them, the point is that the type of hypothesis used matters in the data analysis phase. For those who are familiar with basic statistics, a one-tailed test is appropriate for testing a directional hypothesis. For a nondirectional hypothesis, a two-tailed test is used. (For simplicity, these examples assume two groups—one treatment and one control.) The graphic illustration of that is found in the normal distribution. A nondirectional or hypothesis of differences is tested with a two-tailed t test. Because no direction was specified, the hypothesis is supported if the t value falls in either end of the normal distribution, marked by the blue shaded areas in figure 3.2.

A directional hypothesis is only supported if the t value falls in the end of the normal distribution that is specified in the hypothesis. In this example, the hypothesis is that photographs would lead to *higher* levels of moral judgment, so the t value must fall in the blue shaded area to be significant, represented in figure 3.3. Practically speaking, if significance exists, it is easier to find it with a directional hypothesis; however, if the direction predicted is wrong, then a



^{iv}Other uses of research questions will be covered later in this chapter.

FIGURE 3.2 ■ NONDIRECTIONAL HYPOTHESIS**FIGURE 3.3** ■ DIRECTIONAL HYPOTHESIS

statistical test may find significance but in the opposite direction that was predicted, in which case the hypothesis is not supported.

Hypothesis Writing Formula

Now that this textbook has covered when and why to propose a hypothesis that specifies a direction or not, I offer a basic formula to write clear and focused hypotheses. For each hypothesis, the formula includes:

- Prediction of either a statistically significant difference or direction of that difference (shown in underlines as follows).
- The independent variable (shown in **bold**).
- The dependent variable, stated in conceptual terms (shown in double underline).

For experiments that use two groups, a treatment and control, hypotheses use the term *between*. Future tense is also preferred. Written out as if we were filling in the blanks, this looks like:

- There will be a significant difference in credibility ratings (the DV) between subjects who see the story on **Twitter** and those who see it on a **website** (the IV).⁵⁸
- There will be a significant difference in importance ratings (the DV) between subjects who see **long or short stories** (the IV).⁵⁹
- Audiences exposed to the **advocacy frame** (IV) will rank crime as significantly different in importance (DV) than those exposed to the **objective stories** (IV).⁶⁰

These are all nondirectional hypotheses, as they specify a difference but do not say whether that difference will be higher or lower, large or small. They could be written directionally, for example:

- Credibility ratings (the DV) will be significantly higher for subjects who see the story on **Twitter** than for those who see it on a **website** (the IV).
- Importance ratings (the DV) will be significantly lower for subjects who see a **short story** than for those who see a **long story** (the IV).
- Audiences exposed to the **advocacy frame** (IV) will rank crime as significantly more important (DV) than those exposed to the **objective frame** (IV).

If an experiment involves three or more groups—for example, two treatments and a control group—the term *among* is used instead of *between*, like this:

- There will be a significant difference in credibility ratings (the DV) among subjects who see the story on **Twitter**, a **website**, or in a **newspaper** (the IV).

Notice this one has simply changed the IV to include three conditions—Twitter, a website, and a newspaper—instead of two.

In all these examples of basic hypotheses, each is precisely worded. They use clear, simple, single cause-and-effect predictions. The same words are used throughout rather than synonymous terms. These should be the same terms used in the literature review. Each predicts a clear causal relationship between specific IVs and DVs. Directional hypotheses are preferred if they can be supported by theory and/or prior evidence.

If an experiment includes covariates or statistical controls for individual differences that have been shown to affect the DV but are not equivalently distributed during random assignment, the phrase *controlling for age, education, and gender*, or whatever the variables are, is added to the end of the hypothesis:

H1: There will be a statistically significant difference in credibility ratings among subjects who see the story on Twitter, a website, or a newspaper, controlling for age, education, and gender.

Hypotheses With More Than One IV

So far in this book, we have talked about manipulating one independent variable at a time—for example, the format of the message in the preceding hypothesis is whether subjects see it on Twitter, a website, or in a newspaper. In the previous example, the independent variable is the frame: advocacy or objective. These are known as *single-factor designs* and will be explained in detail in a later chapter. In reality, many experiments manipulate more than one independent variable at a time—for example, the format of the message *and* its frame, in the same study. When more than one independent variable, or **factor**, is manipulated, separate hypotheses should be written for each factor, also called a *main effect*, and for the **interaction** of the two factors or the effects of the two factors considered together. So, for example, in an experiment that manipulates the message format (Twitter, website) and the frame (advocacy, objective), there should be two hypotheses, one for each independent variable or factor's main effect:

H1: Credibility scores will be significantly higher for subjects who read stories on Twitter than on a website.

H2: Credibility scores will be significantly higher for subjects who read objectively framed stories than advocacy framed stories.

Finally, there should be either a hypothesis or a research question about the interaction between the format and the frame, such as:

H3: Credibility scores will be higher for subjects who read objectively framed stories on a website than for any of the other combinations of message format and frame;

OR

RQ1: Is there an interaction between message format and frame?

Naturally, the choice of whether to make a prediction or ask a research question depends on theory and evidence; use a hypothesis whenever possible. This next example of an interaction hypothesis shows the two experimental manipulations or IVs—disgust and harm—interacting with subjects' preexisting political views. The author predicts different effects of the IVs depending on whether the subject supports or opposes a political practice:

“H4: Incidental disgust and harm associations will increase moral conviction and lead to a harsher moral judgment among opponents of a political practice but lead to the opposite effect among supporters.”⁶¹

These hypotheses for studies with more than one independent variable, or factor, also build in a logical progression, from one IV to another, then to the interaction of the two IVs. If intervening variables are proposed, then they too are taken one at a time in a logical order. To think more clearly about these more complicated hypotheses, it may help to create a table listing the independent variables or factors and their levels, the intervening variables, and dependent variables. Use these to visualize the hypotheses, as shown in How To Do It box 3.4.

Here is an example of a hypothesis that includes an intervening variable: the concept of powerfulness. This experiment studies spokespersons who respond to an organizational crisis of contaminated food. The IV in this experiment is voice pitch, with the two levels being high pitch or low pitch. The intervening variable is powerfulness, which has been perceived in low-pitched voices. Thus, the hypothesis says:

“A lowered voice pitch will result in greater perceptions of competence compared to a raised voice pitch due to an intermediate effect of perceived powerfulness (H2).”⁶²

There are many more variations on hypotheses in published journal articles; this is meant to introduce the beginner to the basics of hypothesis writing. As you read published experiments and conduct your own, adapt your hypothesis writing to the needs of each study.

HOW TO DO IT 3.4

Hypothesis Writing Table

To more easily diagram out the steps for writing hypotheses, it may be helpful to visualize the variables with a table, as follows. List the independent variables or factors in a column on the left, with the levels of each. List the dependent variables in the column on the right. If there are any intervening variables or causal mechanisms, list them in the middle column. Then, draw lines to demonstrate the process and predictions that will become the hypotheses. The tables that follow are simplified but can be expanded to accommodate as many factors, intervening variables, and dependent variables as needed.

For an Interaction Effect When There Are Two or More Factors

Independent Variables or Factors & Levels	Intervening Variables or Causal Mechanisms	Dependent Variables
Platform <i>Twitter</i> <i>Website</i>	H1	Credibility
Frame <i>Advocacy</i> <i>Objective</i>	H3	

- H1: Subjects who read stories on Twitter will have significantly higher credibility scores than subjects who read stories on a website. (Main effect of Platform factor.)
- H2: Subjects who read objectively framed stories will have significantly higher credibility scores than subjects who read advocacy framed stories. (Main effect of Frame factor.)
- H3: Credibility scores will be higher for subjects who read objectively framed stories on a website than for any of the other combinations of message format and frame. (Interaction of Platform and Frame factors.)

For Hypotheses With Causal Mechanisms That Build on Each Other

Independent Variables or Factors & Levels	Intervening Variables or Causal Mechanisms	Dependent Variables
Photographs <i>See</i> <i>Do Not See</i>	H1 Elaboration about Stakeholders H2 H3	Ethical Reasoning

(Continued)

(Continued)

H1: Participants who see photographs will have significantly higher levels of elaboration about stakeholders than those who do not. (Factor leads to causal mechanism.)

H2: More elaboration about stakeholders will be significantly associated with higher ethical reasoning. (Causal mechanism leads to DV.)

H3: Ethical reasoning will be significantly higher for participants who see photographs than for those who do not. (Factor leads to DV.)

One final piece of advice on writing hypotheses is to think about the importance of each one. Ask if the hypotheses are interesting, meaningful, or important. Are they self-evident or already well tested? Try not to propose what I call “duh hypotheses”—that is, those whose answer is obvious. For example, the basic agenda-setting hypothesis that the issues the media cover most are the same ones that the public says are most important is well tested. There is simply no earth-shattering revelation to be had by testing this as if it were the first time anyone had asked it. Many agenda-setting studies never ask it. When they do, it is frequently used to establish that a media agenda existed as a basis for more specific predictions, or to see if untested contingent conditions make the outcome different than what has been shown repeatedly for forty-plus years.

Also, ask what it would mean if one of these three things happened:

- What would it mean if the hypothesis was supported?
- What would it mean if it was not? (No significance)
- For directional hypotheses, what would it mean if significance was found in the opposite direction than predicted?

This exercise is designed to help you think about all the possible outcomes and what it would mean if each occurred. This will help you think about the meaningfulness of your study and also how you might explain the results in the discussion section. If you cannot imagine the conclusions under any of these situations, or they seem obvious or unimportant, it is time to rethink the study. Also, think about what might be included in the study as evidence to back up the explanations. For example, if data fail to support the hypothesis or find significance in the opposite direction than predicted, why might that have happened? This exercise also addresses a phenomenon I find frequently among novice

experimentalists—feeling they have “failed” if their hypotheses are not supported. Even in studies that fail to support the hypotheses, something valuable has been learned. In some cases, not supporting a hypothesis is more interesting than supporting it. This, after all, is the purpose of conducting research in the first place, not merely to support all hypotheses. As Campbell and Stanley say, “The task of theory-testing . . . is therefore predominantly one of rejecting inadequate hypotheses.”⁶³ Rest assured, failing to support the hypotheses does not make a study unpublishable.

Research Questions

In addition to being used to probe interactions between two independent variables, research questions are sometimes found for main effects. These are made when there is not enough evidence or theory to support a prediction, and also when models are being tested. Writing research questions for experiments is no different than for other methods. In addition to the formal research question style shown earlier, research questions can also be written in narrative style, for example:

In sum, there are four primary questions addressed by this study. First, do the two teaching interventions differ in their overall effectiveness with respect to CVS learning? Second, do the two teaching interventions differ in the degree to which learning transfers to new domains? Third, do students learn different things from the two interventions? Specifically, do they develop a better understanding of the need to control irrelevant variables from the intervention based on invalid designs? Fourth, do the two teaching interventions differ in their relative effectiveness for students in different learning environments (i.e., higher vs. lower achieving schools)?⁶⁴

The process of designing an experiment does not necessarily work in the same order as the chapters in this book. Sometimes, researchers need to revisit something they have already done. Writing good, clear hypotheses is one of those things. In addition, it is usually necessary to write, rewrite, edit, rewrite, edit, and rewrite again before being satisfied with the precision and clarity of the hypotheses.

The next chapter will discuss experimental designs, including the classic versions of the true laboratory experiments this book focuses on, as well as quasi, natural, and field experiments. As with hypothesis writing, once you have thought through your design, it might be necessary to go back and rework the hypotheses and research questions. This is normal, and research is messy.

Common Mistakes

- Not knowing the relevant literature thoroughly enough
- Orphan Variables—not having all the variables conceptually defined in the hypotheses or research questions, and operationally defined in the methods sections
- Not explaining the posited relations among variables and how they relate to theory

Test Your Knowledge

1. Which of these is an example of a conceptual definition?
 - a. Authoritarianism is defined as a personality trait strongly related to aversion to uncertainty and threat sensitivity.
 - b. Authoritarianism is measured by valuing respect for elders, having good manners, being obedient, and being well behaved.
 - c. Authoritarianism is measured on a 7-point scale.
 - d. Authoritarianism is defined differently by various researchers.
2. Which type of hypothesis does the following represent?
 “Sources with high levels of expertise will positively affect perceptions of credibility.”
 - a. Directional hypothesis
 - b. Nondirectional hypothesis
 - c. Interaction hypothesis
 - d. Null hypothesis
3. Whether you make a directional or nondirectional hypothesis is determined by how much risk you are willing to take that your hypothesis will be supported or not.
 - a. True
 - b. False
4. When you have more than one hypothesis, always start with the one that represents the final outcome of the process you are testing.
 - a. True
 - b. False

5. The purpose of a literature review is to _____.
 - a. Develop theoretical contributions
 - b. Explain how variables are measured
 - c. Provide enough detail that other researchers can replicate the study
 - d. Build upon what has been discovered previously
6. The idea for an experiment should come from:
 - a. Hunches and conventional wisdom
 - b. A process of trial and error
 - c. Theory and existing evidence
 - d. None of these
7. Hypotheses should _____.
 - a. Be described as proven or not
 - b. Be stated in terms of operationalizations
 - c. Go in the methods section
 - d. Build logically upon each other
8. Research questions should be used instead of hypotheses when there is not enough evidence to make a prediction
 - a. True
 - b. False
9. "Journalists who see photographs will use significantly different levels of moral judgment than journalists who do not see photographs." This statement represents which of the following?
 - a. A directional hypothesis
 - b. A nondirectional hypothesis
 - c. A research question
 - d. A null hypothesis
10. Whether to make a directional or nondirectional hypothesis depends upon _____.
 - a. Theory
 - b. Evidence
 - c. The statistical test being used
 - d. Both A and B

For Review-No Commercial Use(2023)

Answers:

- | | | | |
|------|------|------|-------|
| 1. a | 4. b | 7. d | 9. b |
| 2. a | 5. d | 8. a | 10. d |
| 3. b | 6. c | | |

Application Exercises

- Choose one of your three ideas from the assignments in chapter 1 and add to it at least five pages that review the theory it will develop, test, or extend, and the literature related to it. Use at least twenty-five articles. Be sure to analyze and synthesize the literature, not just summarize it, and to connect the literature to your study.
- Write two to three clear and concise hypotheses to go with the experiment you are developing. Use the formula in this chapter. If you have more than one factor (IV) in your experiment, write a hypothesis for the main effect of each factor and a hypothesis for the interaction effect. Have two colleagues read it to see if it is clear.

For Review-No Commercial Use(2023)

Suggested Readings

From the University of Washington's Psychology Writing Center, this white paper on how to write a literature review:

https://depts.washington.edu/psych/files/writing_center/litrev.pdf

Shoemaker, P., J. Tankard, and D. Lasorsa. 2003. *How to Build Social Science Theories*. Thousand Oaks, CA: Sage.

Chapter 2, "Manuscript Structure and Content," in the *Publication Manual of the American Psychological Association*, 6th ed. (2010). Washington, DC: American Psychological Association.

Chapters 2, 3, and 4 in Bausell, R. Barker. 1994. *Conducting Meaningful Experiments: 40 Steps to Becoming a Scientist*. Thousand Oaks, CA: Sage.

Notes

- | | |
|---|--|
| 1. R. Barker Bausell, <i>Conducting Meaningful Experiments: 40 Steps to Becoming a Scientist</i> (Thousand Oaks, CA: Sage, 1994), 32. | 2. Thomas S. Kuhn, <i>The Structure of Scientific Revolutions</i> , 3rd. ed. (Chicago: University of Chicago Press, 1996); Earl Babbie, <i>The Practice of</i> |
|---|--|

- Social Research*, 11th ed. (Belmont, CA: Thompson-Wadsworth, 2007).
3. H. W. Turnbull, ed. *The Correspondence of Isaac Newton: 1661–1675, Volume 1* (London: The Royal Society at the University Press, 1959).
 4. Esther Thorson, Robert H. Wicks, and Glenn Leshner, “Experimental Methodology in Journalism and Mass Communication Research,” *Journalism and Mass Communication Quarterly* 89, no. 1 (2012): 112–124.
 5. Rebecca B. Morton and Kenneth C. Williams, *Experimental Political Science and the Study of Causality: From Nature to the Lab* (New York: Cambridge University Press, 2010).
 6. Paul D. Leedy and Jeanne Ellis Ormrod, *Practical Research: Planning and Design* (Boston: Pearson Education Inc., 2010), 5.
 7. Kurt Lewin, *Field Theory in Social Science: Selected Theoretical Papers by Kurt Lewin* (London: Tavistock, 1952), 169.
 8. Leedy and Ormrod, *Practical Research*.
 9. S. Iyengar, *Is Anyone Responsible?: How Television Frames Political News* (Chicago: University of Chicago Press, 1991).
 10. L. R. Franklin, “Exploratory Experiments,” *Philosophy of Science* 72, no. 5 (2005): 888–899; C. Kenneth Waters, “The Nature and Context of Exploratory Experimentation: An Introduction to Three Case Studies of Exploratory Research,” *History and Philosophy of the Life Sciences* 29, no. 3 (2007): 275–284.
 11. Franklin, “Exploratory Experiments”; Waters, “The Nature and Context of Exploratory Experimentation.”
 12. Franklin, “Exploratory Experiments.”
 13. Ibid.
 14. Babbie, *The Practice of Social Research*.
 15. Bausell, *Conducting Meaningful Experiments*.
 16. Ibid.
 17. Renita Coleman, “The Effect of Visuals on Ethical Reasoning: What’s a Photograph Worth to Journalists Making Moral Decisions?” *Journalism and Mass Communication Quarterly* 83, no. 4 (2006): 835–850.
 18. H. B. Brosius, “The Effects of Emotional Pictures in Television News,” *Communication Research* 20, no. 1 (1993): 105–124; David Domke, David Perlmutter, and Meg Spratt, “The Primes of Our Times? An Examination of the ‘Power’ of Visual Images,” *Journalism* 3, no. 2 (2002): 131–159; A. Friedman, “Framing Pictures: The Role of Knowledge in Automatized Encoding and Memory for Gist,” *Journal of Experimental Psychology: General* 108 (1979): 316–335; Doris Graber, “Seeing in Remembering: How Visuals Contribute to Learning From Television News,” *Journal of Communication* 40, no. 3 (Summer 1990): 134–155; George E. Marcus, W. Russell Neuman, and Michael MacKuen, *Affective Intelligence and Political Judgment* (Chicago: University of Chicago Press, 2000); Richard E. Petty and John T. Cacioppo, *Communication and Persuasion: Central and Peripheral Routes to Attitude Change* (New York: Springer-Verlag, 1986).
 19. Elinor Amit and Joshua D. Greene, “You See, the Ends Don’t Justify the Means: Visual Imagery and Moral Judgment,” *Psychological Science* 23, no. 8 (2012): 861–868; J. D. Greene et al., “The Neural Bases of Cognitive Conflict and Control in Moral Judgment,” *Neuron* 44 (2004): 389–400; J. D. Greene et al., “An fMRI Investigation of Emotional Engagement in Moral Judgment,” *Science* 293 (2001): 2015–2019; Joshua D. Greene and Jonathan Haidt, “How (and Where) Does Moral Judgment Work?” *Trends in Cognitive Sciences* 6, no. 12 (2002): 517–523; D. Kahneman and S. Fredrick, “Representativeness Revisited: Attribute Substitution in Intuitive Judgment,” in *Heuristics and Biases*, ed. T. Gilovich, D. Griffin, and D. Kahneman (New York: Cambridge University Press, 2002), 49–81; Emma Rodero, “See It on a Radio Story: Sound Effects and Shots to Evoked Imagery and Attention on Audio Fiction,” *Communication Research* 39, no. 4 (2012): 458–479; Adam B. Moore, Brian A. Clark, and Michael J. Kane, “Who Shalt Not Kill? Individual Differences in Working Memory Capacity, Executive Control, and Moral Judgment,” *Psychological Science* (0956-7976) 19, no. 6 (2008): 549–557.
 20. Vicki Goldberg, *The Power of Photography: How Photographs Changed Our Lives* (New York: Abbeville, 1991).
 21. Bausell, *Conducting Meaningful Experiments*, 21.
 22. H. Denis Wu and Renita Coleman, “Advancing Agenda-Setting Theory: The Comparative Strength and New Contingent Conditions of the Two Levels of Agenda-Setting Effects,” *Journalism and Mass*

- Communication Quarterly* 86, no. 4 (Winter 2009): 775–789.
23. James Potter, *On Media Violence* (Thousand Oaks, CA: Sage, 1999).
 24. Thorson, Wicks, and Leshner, “Experimental Methodology.”
 25. Pamela Shoemaker, James W. Tankard, and Dominick Lasorsa, *How to Build Social Science Theories* (Thousand Oaks, CA: Sage, 2003), 11.
 26. *Ibid.*, 11.
 27. George Gerbner et al., “Living with Television: The Dynamics of the Cultivation Process,” in *Perspectives on Media Effects*, ed. Jennings Bryant and C. Zillmann (Hillsdale, NJ: Erlbaum, 1986), 17–40; George Gerbner et al., “Growing Up with Television: The Cultivation Perspective,” in *Media Effects: Advances in Theory and Research*, ed. J. Bryant and C. Zillmann (Hillsdale, NJ: Erlbaum, 1994), 7–14.
 28. *Ibid.*
 29. Coleman, “The Effect of Visuals on Ethical Reasoning.”
 30. Wu and Coleman, “Advancing Agenda-Setting Theory,” 778.
 31. *Ibid.*, 777.
 32. Seung-A Annie Jin and Joe Phua, “Following Celebrities’ Tweets About Brands: The Impact of Twitter-Based Electronic Word-of-Mouth on Consumers’ Source Credibility Perception, Buying Intention, and Social Identification with Celebrities,” *Journal of Advertising* 43, no. 2 (2014): 183.
 33. Nan-Hyun Um and Wei-Na Lee, “Does Culture Influence How Consumers Process Negative Celebrity Information? Impact of Culture in Evaluation of Negative Celebrity Information,” *Asian Journal of Communication* 25, no. 3 (2015): 329.
 34. Elena Novak and Janet Tassell, “Using Video Game Play to Improve Education-Majors’ Mathematical Performance: An Experimental Study,” *Computers in Human Behavior* 53 (2015): 125.
 35. Horacio Alvarez-Marinelli et al., “Computer Assisted English Language Learning in Costa Rican Elementary Schools: An Experimental Study,” *Computer Assisted Language Learning* 29, no. 1 (2016): 105.
 36. Ryan Oprea, Bart J. Wilson, and Arthur Zillante, “War of Attrition: Evidence from a Laboratory Experiment on Market Exit,” *Economic Inquiry* 51, no. 4 (2013): 2019.
 37. Miglena Mantcheva Sternadori and Esther Thorson, “Anonymous Sources Harm Credibility of All Stories,” *Newspaper Research Journal* 30, no. 4 (2009): 56.
 38. James R. Rest, Lynne Edwards, and Stephen J. Thoma, “Designing and Validating a Measure of Moral Judgment: Stage Preference and Stage Consistency Approaches,” *Journal of Educational Psychology* 89, no. 1 (March 1997): 5–28.
 39. Sternadori and Thorson, “Anonymous Sources Harm Credibility of All Stories,” 56.
 40. Kristin Bulkow, Juliane Urban, and Wolfgang Schweiger, “The Duality of Agenda-Setting: The Role of Information Processing,” *International Journal of Public Opinion Research* 25, no. 1 (Spring 2013): 43–63.
 41. *Ibid.*, 52.
 42. Tara Halle, “A Cautionary Tale: Have You Checked That Citation?” *Covering Health* (2017); Lisa Marriott, “Using Student Subjects in Experimental Research: A Challenge to the Practice of Using Students as a Proxy for Taxpayers,” *International Journal of Social Research Methodology* 17, no. 5 (2014): 503–26.
 43. Mirae Kim and Gregg G. Van Ryzin, “Impact of Government Funding on Donations to Arts Organizations: A Survey Experiment,” *Nonprofit and Voluntary Sector Quarterly* 43, no. 5 (2014): 913.
 44. Thomas Zerback, Thomas Koch, and Benjamin Kramer, “Thinking of Others: Effects of Implicit and Explicit Media Cues on Climate of Opinion Perceptions,” *Journalism and Mass Communication Quarterly* 92, no. 2 (2015): 421–443.
 45. *Ibid.*, 423.
 46. *Ibid.*, 424.
 47. *Ibid.*, 425.
 48. Tianguang Meng, Jennifer Pan, and Ping Yang, “Conditional Receptivity to Citizen Participation,” *Comparative Political Studies* (2014): 8.
 49. Coleman, “The Effect of Visuals on Ethical Reasoning,” 839.
 50. Diana C. Mutz, “Effects of ‘In-Your-Face’ Television Discourse on Perceptions of a Legitimate Opposition,” *American Political Science Review* 101, no. 4 (2007): 624.
 51. D. T. Campbell and J. C. Stanley, *Experimental and Quasi-Experimental Designs for Research*. (Chicago: Rand McNally, 1963), 35.

52. Thorson, Wicks, and Leshner, "Experimental Methodology."
53. Ibid.
54. Y. J. Cho and E. J. Ringquist, "Managerial Trustworthiness and Organizational Outcomes," *Journal of Public Administration Research and Theory* 21 (2011): 53–54.
55. Mogens Jin Pedersen and Justin M. Stritch, "Internal Management and Perceived Managerial Trustworthiness," *American Review of Public Administration* (2016): 5.
56. Ibid.
57. Thorson, Wicks, and Leshner, "Experimental Methodology."
58. Adapted from Mike Schmierbach and Anne Oeldorf-Hirsch, "A Little Bird Told Me, So I Didn't Believe It: Twitter, Credibility, and Issue Perceptions," *Communication Quarterly* 60, no. 3 (July-August 2012): 317–337.
59. Ibid.
60. Adapted from Sean Aday, "The Framesetting Effects of News: An Experimental Test of Advocacy Versus Objectivist Frames," *Journalism and Mass Communication Quarterly* 83, no. 4 (Winter 2006): 767–784.
61. Pazit Ben-Nun Bloom, "Disgust, Harm and Morality in Politics," *Political Psychology* 35, no. 4 (2014): 500.
62. An-Sofie Claeys and Verolien Cauberghe, "Keeping Control: The Importance of Nonverbal Expressions of Power by Organizational Spokespersons in Time of Crisis," *Journal of Communication* 64 (2014): 1162.
63. Campbell and Stanley, *Experimental and Quasi-Experimental Designs for Research*, 35.
64. Robert F. Lorch et al., "Using Valid and Invalid Experimental Designs to Teach the Control of Variables Strategy in Higher and Lower Achieving Classrooms," *Journal of Educational Psychology* 106, no. 1 (February 2014): 18–35.

For Review-No Commercial Use(2023)



For Review-No Commercial Use(2023)



TYPES OF EXPERIMENTS

Truth has nothing to do with the conclusion, and everything to do with the methodology.

—Stefan Molyneux

LEARNING OBJECTIVES

For Review-No Commercial Use(2023)

- Summarize the different types of experiments using Campbell and Stanley's typology.
- Recommend when to use each of the three true experimental designs.
- Critique the strengths and weaknesses of designs with pretests.
- Describe quasi, natural, and field experiments.
- Explain how quasi, natural, and field experiments differ from each other and from true experiments.

With the literature review written or at least in draft form, as discussed in the previous chapter, the next step is to decide what basic type of experiment to conductⁱ—that is, the methodology or system of the experiment itself referred to in this chapter's opening quote. One authoritative word on this topic is the classic *Experimental and Quasi-Experimental Designs for Research* by Donald T. Campbell and Julian C. Stanley.¹

ⁱNormally, experiments are not designed in such a linear fashion. Typically, researchers think about all the issues covered in this book simultaneously.

This work is only seventy-one pages long without the references. It started life as a chapter in a research handbook in 1963 but is so popular that it continues to be published as a monograph. It describes sixteen different ways to do an experiment, all still valid today, although some designs are more popular than others. It was written for the teaching field, so examples are from education research. Next is a summary of Campbell and Stanley's typology of experimental designs, highlighting six designs popular in social science along with a critique of each. Most experiments actually conducted today are slight variations of the designs described here. In addition to these designs in Campbell and Stanley's typology, this chapter will also briefly describe quasi experiments, natural experiments, and field experiments.

CAMPBELL AND STANLEY'S TYPOLOGY OF EXPERIMENTS

Campbell and Stanley use graphic shorthand to describe experiments that should be familiar to football fans but have different meanings here.

For Review-No Commercial Use(2023)
 X = an exposure, treatment, manipulation, or intervention. Usually an independent variable (IV).

O = an observation or measurement of an outcome variable. Usually a dependent variable (DV). These are the data that are recorded, either by the researcher, with technical instruments such as a heart rate monitor, or a self-report by the subject. If observations are made or data recorded more than once, that is indicated by a subscript—O₁, O₂—meaning the first and second observation.

If there is a space (____) and no X or O, that means no treatment was given or no observation made.

R = random assignment of subjects to conditions. This will be discussed more in chapter 7. Basically, this is the element that qualifies an experiment as a “true” experiment versus a “quasi” experiment.

The first three design types are classified as “pre-experimental designs” and are “of almost no scientific value,” according to Campbell and Stanley.² These designs tend to be reported as “exploratory” in articles I review, but “pre-experimental” is a more accurate term. They also qualify as quasi experimental designs because of the lack of random assignment. Here, I briefly review them but recommend using other designs as described.

Three Pre-Experimental Designs

The One-Shot Case Study

This can sometimes be seen reported in news stories when someone wants to attribute a cause to some effect—for example, how school absentee rates went down compared to previous years after some intervention. It might be something like a school that implemented a text messaging system that pinged students in time to get them to school by 9 a.m. This is graphically represented as:

— X O

The text messaging system was the intervention (X) and absentee rates the observation (O). There are several problems with this, however—among them that the students who were observed were not randomly sampled, there is no control group of students who did not get the text messages to compare against, and there were no controls in place that would rule out alternative explanations. For example, some students' parents could have started making waffles for breakfast, new construction outside others' windows woke them up early, some could have gotten new cars and were excited to get to school to show them off, and maybe others saw a presentation on the importance of college and suddenly figured they needed to get to school in order to be able to support themselves. All these things—no controls, no randomization—could have explained the lower absentee rates, not just the text messaging system. Because of these problems, the one-shot case study is not a good design for a true social science experiment.

One-Group Pretest–Posttest Design

Slightly better than the one-shot case study is when an observation or measurement is added before the treatment is given, called a **pretest**. The same observation or measurement is given after the intervention (the posttest), so researchers have a baseline measure to compare any changes against.

O₁ X O₂

For example, researchers might measure how much math a student knows in the pretest, then give the intervention, such as computer-aided tutoring, and then test the students again to see how much they learned. The difference between the pretest and posttest scores represents the effect. Sometimes, however, adding a pretest does not necessarily make things better, because people tend to learn how to do better on a test after they have taken it once or when they realize they are being watched. For example, those who take IQ tests more than once have been shown to get smarter on the second try.³ In research, this effect of testing is called “test, retest gain.”⁴ There could be other explanations as well. Reactivity is the idea that measuring something changes it; just by knowing they are being observed, people tend to do better⁵ (see More About box 4.1).

MORE ABOUT . . . BOX 4.1

The Hawthorne Effect



Western Electric Company

Hawthorne, Illinois, Works of the Western Electric Company, 1925

The Hawthorne effect is a specific type of reactivity, the idea that people react to being observed and change what they do. It is important in research because some people will not give true answers if they know they are participating in a study. This effect is not necessarily intentional; sometimes, people simply change without realizing it. The Hawthorne effect is the popular, although some say inappropriate, name for a demand characteristic that arises when subjects of study change their behavior because they know they are being observed.⁶

It developed from studies of worker productivity commissioned by the Hawthorne Works, a Western Electric factory outside Chicago from 1924 to 1932.⁷ They are also known as the illumination studies because higher and lower levels of lighting were tested for changes in worker productivity; in reality, many other variables were also studied, including work hours, break times, the cleanliness of floors and work stations, among others.⁸ The truncated version of the study results is that when the researchers changed the workers' lighting and break times, their productivity improved. In spite of the changes that led to improvements staying in place, worker productivity fell when the study was over. The original conclusion that paying attention to workers would result in greater productivity was later reinterpreted to say that people change their behavior when they know they are being observed.⁹

Not all studies show a Hawthorne effect,¹⁰ and research still investigates it today,¹¹ especially in the health sciences, human relations, and organizational behavior.¹² The illumination studies, for there were many over several years, are more complex than usually presented. For example, it is a myth that improvement was continuous, and there were potential confounding variables such as learning, feedback, and incentive pay. Thus, the term *Hawthorne effect* to describe reactivity is diminishing in use.¹³

A similar phenomenon is known as the **demand effect**, or demand characteristics—the idea that some people change their answers or behavior in order to please an experimenter.¹⁴ In studying the experimental situation, Martin Orne discovered that study subjects tried to guess the purpose and altered their behavior to fit their interpretations.¹⁵ Similar phenomena include the *halo effect* or *social desirability effect*, where study participants try to portray themselves in a positive light.¹⁶

The Hawthorne studies make another important contribution to experimental design in that they show the importance of manipulation checks,¹⁷ which will be covered in chapter 9. In the Hawthorne studies, it was not so much the manipulations that had an effect but the workers' interpretation of them, and understanding subjects' interpretations is the purpose of a manipulation check.

There could also be a change in the measurement standards if observers are recording the data. There is also a phenomenon known as *regression to the mean*, whereby those who score extremely well or poorly tend to go back toward the middle the next time they are tested.¹⁸ Moreover, if there is a time gap between the first and second observation, something else could have happened. For example, if the observation being recorded is a person's level of fear about flying, and if an airplane crashes and is reported in the news between O_1 and O_2 , that could change the outcome.

Static Group Comparison

In this type of design, a second group that has not received the treatment has been added (O_2).

$$\begin{array}{l} X \quad O_1 \\ \text{—} \quad O_2 \end{array}$$

An example would be comparing people who saw the presidential candidates' debate versus those who did not, or comparing students who got antibullying training to those who did not. There are also problems with this, including that there is no way to tell if the people in the two groups are the same on important individual characteristics—for example, perhaps there were more Republicans in the debate-watching group than the nonwatching group, or more aggressive students in one group than the other. This will be discussed in greater detail in chapter 7 on random assignment. People's political identification and students' innate aggressive tendencies could affect the outcome.

As I do not recommend these three designs, they will not be discussed in detail regarding their strengths and weaknesses. Instead, this chapter will concentrate on the next three, which are true experimental designs that I do recommend.

Three True Experimental Designs

Pretest–Posttest Control Group

This design is one of the most used in social science. It adds the crucial feature of randomly assigning subjects to either the treatment or control conditions, thus making sure the groups are equal on important characteristics that could otherwise cause any changes.

$$\begin{array}{l} R \quad O_1 \quad X \quad O_2 \\ R \quad O_3 \quad \text{—} \quad O_4 \end{array}$$

The drawback of this design is the same as in the one-group pretest–posttest design—that is, being observed or measured twice may cause changes in the subjects’ performance, attitudes, or whatever else the outcome is. Because this threat has been shown to be so prevalent, the popularity of pretests has been declining, and pretests are actually not essential to true experimental designs.¹⁹

Solomon Four-Group Design

It is common for the Solomon four-group design to be described as the gold standard. And it is. It eliminates all the drawbacks described earlier, plus the researcher can actually tell if there are any effects of testing by having groups where no pretest is given.

R	O ₁	X	O ₂
R	O ₃	—	O ₄
R	—	X	O ₅
R	—	—	O ₆

But it is also time consuming, costly, difficult, and has statistical issues, so relatively few studies actually use it.²⁰ Bausell even calls it “wasteful.”²¹ In this design, there are four groups, with subjects randomly assigned to all of them. There are two control groups consisting of subjects who do not get the treatment or manipulation; they serve as the baseline for comparison with the groups that did get the treatment. One of the treatment groups is given a pretest and one is not, and one of the control groups is given a pretest and the other is not. This allows the researcher to compare not only the differences before and after the treatment, but also to see if the pretest affected the results. This provides what Levy and Ellis call a “defensible response to most rival hypotheses.”²² It also requires many, many more subjects to participate in the experiment, the costs of which sometimes outweigh the benefits. (See Study Spotlight 4.2 for an example of a Solomon four-group design.)

Posttest-Only Control Group Design

The final design in the Campbell and Stanley typology is one that includes no pretests and is the one I use most often. It is also the one that is now most recommended.²³

R	X	O
R	—	O

STUDY SPOTLIGHT 4.2

A Study Using the Solomon Four-Group Design



SAGE Journal Article:
study.sagepub.com/
coleman

Genç, M. 2016. "An Evaluation of the Cooperative Learning Process by Sixth-Grade Students." *Research in Education* 95 (1): 19–32.

This study used a Solomon design to assess the effects of a particular teaching strategy on sixth graders' science knowledge. The teaching strategy, called cooperative learning, has teachers organize students into groups who do research on their own to learn information and solve problems together. Cooperative learning classrooms were the manipulation or treatment group; classrooms as they were organized already, with teachers presenting information to the students in the traditional way, were used as the control groups.

In this Solomon four-group design, two groups of students got the cooperative learning treatment, and two did not, making them the control groups. Pretests were used for one control group and one treatment group; the other groups did not get pretests. The author says that the reason the Solomon design was used was to be able to know if use of pretests caused effects because students could have learned simply by taking the test twice. Students were randomly assigned to all four groups. Here is how it is described:

"The Solomon four-group design is an attempt to eliminate the possible effect of a pretest. It involves random assignment of subjects to four groups, with two of the groups being pretested and two not. One of the pretested groups and one of the unpretested groups is exposed to the experimental treatment. All four groups are then posttested. Although each group is put through post-experimental evaluation, the pre-experimental evaluations are performed in only two groups, one being an experimental group and the other being a control group. The first two groups are treated as the pretest-posttest control group design, and the other two groups are treated as the posttest-control group design" (p. 22).

The study does not go into whether the pretest affected the outcome; however, it can be seen from the mean scores that the groups that received pretests did score higher on the posttests than those that did not get the pretests. No significance tests are reported in this study to determine if the pretest alone had a significant effect, however. The mean posttest scores for those given the pretests were 27.70 (treatment) and 22.50 (control) versus 26.74 (treatment) and 21.91 (control) for students who were not given the pretest. I have reworked the author's tables into the one that follows in order to illustrate the design and also facilitate comparison between posttest scores for the pretested groups compared to the non-pretested groups.

It should be noted that alternative forms of the tests were used for pre- and posttesting rather than identical tests for both occasions, which helps guard against learning from taking the test twice.

The cooperative learning treatment worked, as both pretested and unpretested students in the cooperative learning classrooms performed significantly better on the science tests given at the end of the experiment.

Some social science disciplines use the Solomon design more than others, so it is important to know where your colleagues stand on this.

(Continued)

(Continued)

Means and Standard Deviations for
a Solomon Four-Group Design Experiment

Group	Pretest	Instruction type	Posttest Scores M (SD)
Treatment 1	X	Cooperative learning	27.70 (2.531)
Treatment 2		Cooperative learning	26.74 (1.797)
Control 1	X	Regular curriculum	22.50 (2.219)
Control 2		Regular curriculum	21.91 (2.327)

In a study comparing the pretest–posttest and posttest-only designs, Gorard²⁴ found the results of the posttest-only design to be “less misleading.” Even Campbell and Stanley²⁵ say that pretests are not essential to true experimental designs and explain that they are misunderstood. When the purpose of a pretest is to ensure equivalence of subjects assigned to conditions, random assignment is an adequate precaution, so pretests are not needed. They point out that almost all of Ronald Fisher’s agriculture experiments had no pretests. Pretesting is still preferred in some disciplines but not all, it is important to know the standard in the field. The interaction of a pretest with the treatment is not large in most cases, but in education, psychology, and sociology in particular, the effects are larger and should not be ignored.²⁶ Some studies straddle the line, so to speak, for example by conducting a pilot study that uses a pretest, and if the effects of testing are ruled out, conduct the actual experiment without a pretest.²⁷ (See How To Do It box 4.3 for examples of how to describe a posttest-only control group design.)

These designs represent a basic structure; actual experiments may use slight variations and still be acceptable. So far, this chapter has reviewed three experimental designs that are classified as “true” experiments and three that are “pre-experimental” from Campbell and Stanley’s typology. Campbell and Stanley go on to describe ten other designs they term “quasi experimental.” This book will not go into them in detail but will instead summarize the essential features of a quasi experiment and refer readers to the Campbell and Stanley book for more details.

QUASI EXPERIMENTS

The key difference between true and quasi experiments is that in quasi experiments, subjects are not randomly assigned to conditions, the groups may not necessarily be