

# Laboratory Experiments in the Social Sciences

*We dedicate this book to the memory of  
Elizabeth G. Cohen and Bernard P. Cohen,  
social scientists who matched intellectual  
passion with logical precision*

# Laboratory Experiments in the Social Sciences

---

Second Edition

Edited by  
**Murray Webster, Jr. and Jane Sell**



AMSTERDAM • BOSTON • HEIDELBERG • LONDON

NEW YORK • OXFORD • PARIS • SAN DIEGO

SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO

Academic Press is an imprint of Elsevier



*We dedicate this book to the memory of  
Elizabeth G. Cohen and Bernard P. Cohen,  
social scientists who matched intellectual  
passion with logical precision*

Academic Press is an imprint of Elsevier  
32 Jamestown Road, London NW1 7BY, UK  
225 Wyman Street, Waltham, MA 02451, USA  
525 B Street, Suite 1800, San Diego, CA 92101-4495, USA

Second edition

Copyright © 2014, 2007 Elsevier Inc. All rights reserved

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

Permissions may be sought directly from Elsevier's Science & Technology Rights Department in Oxford, UK: phone (+44) (0) 1865 843830; fax (+44) (0) 1865 853333; email: [permissions@elsevier.com](mailto:permissions@elsevier.com). Alternatively, visit the Science and Technology Books website at [www.elsevierdirect.com/rights](http://www.elsevierdirect.com/rights) for further information.

### Notice

No responsibility is assumed by the publisher for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions or ideas contained in the material herein. Because of rapid advances in the medical sciences, in particular, independent verification of diagnoses and drug dosages should be made.

### British Library Cataloguing-in-Publication Data

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

### Library of Congress Cataloging-in-Publication Data

A catalog record for this book is available from the Library of Congress

ISBN : 978-0-12-404681-8

For information on all Academic Press publications  
visit our website at [www.store.elsevier.com](http://www.store.elsevier.com)

Typeset by SPI Global, India

Printed and bound in United States of America

14 15 16 17 18 10 9 8 7 6 5 4 3 2 1



Working together  
to grow libraries in  
developing countries

[www.elsevier.com](http://www.elsevier.com) • [www.bookaid.org](http://www.bookaid.org)

# Preface

This second edition of *Laboratory Experiments in the Social Sciences* demonstrates our commitment to the experimental method as a powerful yet elegant tool for theory testing, development, and cumulation. Although there is an impressive literature on statistical analysis of data from experiments, there is much less available on the design, development, and actual conduct of experiments. Experiments are discussed in most general social science research methods textbooks; however, their logic is not developed there and questions of how to design and conduct experiments generally are omitted. It is not unusual for the experimental method to be mentioned as the “gold standard” for assessment of causality and then to assert without evidence that many, perhaps most, questions of interest to social scientists cannot be addressed through experiments. That misunderstanding is refuted in these chapters, in which the authors show how some of the central questions in sociology, political science, economics, and other fields have benefitted from experimental studies.

We hope that this second edition of our book provides a strong foundation for understanding experimental methods, shows how to conduct all phases of experimental research, and corrects misunderstandings that sometimes surround this method. We recently offered a workshop at the American Sociological Association and found that there were many different kinds of scholars interested in experiments: those who were novices and wanted to learn about the logic of the experimental approach, those who were conducting their own research and had a particular problem with design, and those who had conducted many experiments who wanted to update their knowledge. We found the workshop exhilarating, and it impressed upon us the very different kinds of conceptual approaches, issues, and problems that can be addressed by experiments. We have designed the volume to constitute a useful reference for the researchers we met at the workshop as well as the many others across different disciplines that they represent.

Undergraduate courses, and even more so graduate courses in research methods, deal mostly with methods of analyzing existing data, mostly from large-scale surveys. Such courses are aimed at someone who has already collected data or acquired a large data set from a survey and wants to know how to answer questions from it. What is missing? First, there is nothing on the methodology of generating data; those courses offer little help in learning how to get good experimental data. Researchers who want to collect new data find

few guides, and students who want to evaluate the quality of the data find few criteria they can apply. Second, there is little information on appropriate uses of different methods, such as when to conduct a well-designed survey, when to use structured or unstructured observations, and when to interview. A third topic that is missing is how experiments can be used, even to study phenomena that occur in large groups such as business organizations. Unfortunately, one could conclude from many contemporary courses that although experiments might look good in principle, in actuality few social scientists use them except those interested in very specific individual behavior. These chapters provide an antidote to that misunderstanding as well because many of the lessons and examples come from the authors' extensive experience with experimental research.

Most students learn that experiments offer a high degree of control over independent variables and control over the research setting to facilitate observations of dependent variables. They learn about control conditions and experimental conditions and the importance of random assignment of individuals to conditions. They learn that a well-designed experiment permits high confidence that changes in results observed in dependent variables are associated with (maybe even "caused by") changes in independent variables. So far, so good. At this point, they may be excited about learning how to use this design because of its strengths.

Unfortunately, too often there is more, and some of it can be harmful. Students learn about confounding factors that can mislead investigators who wish to impute causality, such as history, maturation, testing, and many others. Those often are presented as if they are inevitable and make it difficult, if not impossible, to know for sure what happened in a study.<sup>1</sup> To make matters worse, some instructors also show a couple of horrible films depicting unethical social science experiments, and they may talk about the infamous Tuskegee medical experiments.<sup>2</sup> Students can conclude that experiments are complicated and really do not show anything because of confounding factors and also that those who experiment on humans are often unethical.

These are false conclusions. The chapters in this book refute each one of them.

In this second edition of *Laboratory Experiments in the Social Sciences*, the chapter authors confront many of the misunderstandings about experimental

---

1. In 1963, Donald T. Campbell and Julian Stanley published a masterful consideration of potential confounding factors as one chapter in a methodology book. Their chapter is so useful that it was reprinted by itself 2 years later, and it is the standard reference on confounding factors. Oddly overlooked is the fact that Campbell and Stanley showed that a true experiment—using random allocation, a control condition, and one or two experimental conditions—controls *all* the confounding factors they identified. Properly understood, their analysis is a strong argument in favor of using experiments whenever possible.

2. Unethical research of any kind by anyone is unacceptable, as several chapters in this book emphasize.

methods, and they go beyond correcting misunderstandings to offer guidelines for how to do experiments well. In addition, they describe some exemplary research programs that have used these methods.

In Part I of the book, the authors address the philosophical and methodological foundations for experiments. In this part, researchers address what kinds of questions might be answered through experiments, differences among different approaches, and how experiments relate to theory and evidence. These chapters also provide information to address the myriad of theoretical, methodological, and practical considerations of actually doing research. Initial considerations involved in ethics and the role of institutional review boards, training experimenters, and developing the general parameters of the experiment are addressed.

In Part II of the book, different experimental programs are presented that highlight different strategies and different kinds of outcomes. These case studies demonstrate how the power of experimental designs directly feeds into theoretical advancement. These examples reach across different disciplines, including sociology, social psychology, psychology, business, communication, economics, and political science. These chapters illustrate how very different kinds of questions at different levels of analysis can be investigated experimentally.

Part III contains two chapters that provide practical advice about how to write about and design experiments for different kinds of research objectives. There are illustrations of different kinds of research, both applied and basic, and the kinds of considerations in developing, designing, and reporting on the research. In addition, different kinds of funding agencies and how to navigate their requirements successfully are addressed.

This book incorporates the experiences of some of the foremost researchers in experimental social science; the authors bring different insights, innovations, and creativity to the chapters. All of them share a commitment to rigorous experimental methods.

The first edition of the book would not have come into being without the vision of General Editor Scott Bentley from Elsevier and the careful, quick responses of Kathleen Paoni. For this second edition, we are grateful for the advice (and prodding) we received from Nikki Levy, our publisher. Barbara Makinster, the senior editorial project manager, was invaluable in the many decisions and operations to produce this edition; Julia Haynes managed the manuscripts, proofs, and details of production with judgment, tact, and thoughtfulness; Dan Hays provided excellent copyediting and advice on improving style. We thank our mentors who encouraged and nurtured our intellectual interest in experimental methods in graduate school and throughout our careers. We thank our colleagues and our students for challenging and inspiring us. We especially thank the researchers who meet annually at the Group Process meetings, a tradition created by Linda D. Molm. The interaction and intellectual excitement and new ideas at every Group Process meeting renew our energy for our own research and for spreading the word. Early experimental research

clearly documented how the ideas and norms of one generation feed into another generation even when the members of the preceding generation are no longer present. This lasting effect of our intellectual ancestors is clearly demonstrated in the Group Process meetings by the continuing collective commitment to theoretical development, methodological precision, and the integrity of the scientific process.

**Murray Webster, Jr.  
Jane Sell**

# Contributors

**Joseph Berger** (Chapter 12), Department of Sociology, Stanford University, Stanford, CA 94305

**Giovanna Devetag** (Chapter 16), Department of Business and Management, LUISS Guido Carli, Rome, 2 00197 Italy

**James E. Driskell** (Chapter 20), Florida Maxima Corporation, Orlando, FL 32804

**Tripp Driskell** (Chapter 20), Institute for Simulation and Training, University of Central Florida, Orlando FL 32826

**Catherine Eckel** (Chapter 15), Department of Economics, Texas A&M University, College Station, TX 77843

**Martha Foschi** (Chapter 11), Department of Sociology, University of British Columbia, Vancouver, BC, Canada V6T 1Z1

**Karen A. Hegtvedt** (Chapter 2), Department of Sociology, Emory University, Atlanta, GA 30322

**Stuart J. Hysom** (Chapter 7), Department of Sociology, Texas A&M University, College Station, TX 77843

**Will Kalkhoff** (Chapter 5), Department of Sociology, Kent State University, Kent, OH 44242

**Jennifer King** (Chapter 20), U.S. Naval Research Laboratory, Washington, DC 20375

**Kathy J. Kuipers** (Chapter 7), Department of Sociology, University of Montana, Missoula, MT 59812

**Michael K. Lindell** (Chapter 18), Hazard Reduction & Recovery Center, Texas A&M University, College Station, TX 77843

**Michael J. Lovaglia** (Chapter 5), Department of Sociology, University of Iowa, Iowa City, IA 52242

**Rose McDermott** (Chapter 13), Department of Political Science, Brown University, Providence, RI 02912

**Linda D. Molm** (Chapter 9), Department of Sociology, University of Arizona, Tucson, AZ 85721

**Chantrey J. Murphy** (Chapter 17), Department of Sociology, Texas A&M University, College Station, TX 77843

**Leda Nath** (Chapter 5), Department of Sociology, University of Wisconsin–Whitewater, Whitewater, WI 53190

- Andreas Ortmann** (Chapter 16), School of Economics, University of New South Wales, Sydney, NSW 2052, Australia
- Madan Pillutla** (Chapter 19), London Business School, London, NW1 4SA, United Kingdom
- Marko Pitesa** (Chapter 19), University of Maryland, College Park, Maryland, 20742
- Srividya Ramasubramanian** (Chapter 17), Department of Communication, Texas A&M University, College Station, TX 77843
- Bruce Reese** (Chapter 10), Department of Sociology, Texas A&M University, College Station, TX 77843
- Jane Sell** (Chapters 1 and 10), Sociology Department, Texas A&M University, College Station, TX 77843
- Robert K. Shelly** (Chapter 4), Department of Sociology, Ohio University, Athens, OH 45701
- Stefan Thau** (Chapter 19), INSEAD, 138676 Singapore, Singapore
- Shane R. Thye** (Chapter 3), Department of Sociology, University of South Carolina, Columbia, SC 29208
- Lisa Slattery Walker** (Chapter 6), Department of Sociology, University of North Carolina–Charlotte, Charlotte, NC 28223
- Murray Webster, Jr.** (Chapters 1 and 21), Department of Sociology, University of North Carolina–Charlotte, Charlotte, NC 28223
- Rick K. Wilson** (Chapter 14), Department of Political Science, Rice University, Houston, TX 77251
- Reef Youngreen** (Chapter 5), Department of Sociology, University of Massachusetts–Boston, Boston, MA 02125
- Morris Zelditch, Jr.** (Chapter 8), Department of Sociology, Stanford University, Stanford, CA 94305

## Part I

# Designing and Conducting Experiments

This revised and expanded edition of *Laboratory Experiments in the Social Sciences* reflects thorough rewriting of nearly all the chapters. Authors have included new material and examples from recent experiences, and they have rewritten several passages to improve accessibility without sacrificing technical rigor where that is needed. These chapters present the most useful current information about how to conduct high-quality experiments in the social sciences.

The book is divided into three parts or sections. This first Part I contains seven chapters introducing the methods, the philosophical and ethical foundations and some controversies associated with experiments, and the practical concerns that investigators face in developing and implementing experimental designs.

Chapter 1, by the editors, tells what experiments are and what they are not, how the method came into the social sciences, compares experiments to other common methods for collecting and interpreting data, and considers strengths and weaknesses of the method. This chapter describes our understanding of the value of experiments for developing sociological knowledge.

Chapter 2, by Karen A. Hegtvedt, presents information on protecting human participants in experimental research. This chapter distills knowledge from her career conducting experimental studies of justice and emotions, from advising faculty members as Chair of Emory University's Department of Sociology, as editor of the leading sociological journal reporting experimental research, and, most directly relevant, her years of service as a member and Chair of Emory University's institutional review board (IRB) that reviews all procedures and oversees research to ensure the welfare of participants. Every institution conducting research has an IRB, and experimenters must learn to successfully interact with them and to ensure

protection of experimental participants. Hegtvedt describes ethical requirements and principles as they apply to this type of work. She reviews recent work analyzing sources of potential problems in research. She also advises on ways to promote sensitive, ethical, and methodologically sound experimental research and to work collaboratively with an IRB.

Chapter 3, by Shane R. Thye, addresses the skeptical or cynical views some social scientists still have of experimentation and shows that many of those are based on outdated views of experimentation. He describes the most important threats to good experimental design and ways to avoid them or to compensate for uncertainty they can produce. This chapter explains the importance of theoretical foundations for experiments (rather than simply trying to “seek what will happen”), explores conditions for inferring causality from different kinds of evidence, and shows why properly designed experiments offer a strong likelihood of producing an increase in understanding.

Chapter 4, by Robert K. Shelly, describes how he and other skilled experimenters select, train, and manage members of research teams who conduct experiments. The research team is not only the face of the research organization for participants but also the source of information an experimenter uses in deciding success or failure of an experiment and of its underlying theoretical ideas. An experiment is a simplified situation, and so details, including dress and speech habits of experimenters, can have great importance. Shelly likens experiments to theatrical performances in which an alternate reality is created and sustained for a particular purpose. Experiments go beyond theater, however, in that a researcher must consider establishing authority relations and mentoring among the research team, among other considerations.

Chapter 5, by Will Kalkhoff, Reef Youngreen, Leda Nath, and Michael J. Lovaglia, addresses some neglected topics in experimental design: who the participants will be, how to recruit them, and what happens after they volunteer for research. These authors summarize a considerable body of recent research on effects of seemingly small differences in procedures associated with recruiting and working with participants. They consider the benefits and costs to participants, and they offer suggestions for maximizing the former as part of good research design. The authors describe the many steps in finding and contacting reliable sources of participants, deciding which of them are suitable for an experimental project, record keeping, and interpersonal relations with participants.

Chapter 6, by Lisa Slattery Walker, is designed to help a research program move from theoretical ideas to experimental research. In other words, once you have some good research questions, how do you develop an experimental design to answer them? This chapter includes a wide range

of suggestions and topics that a researcher must address in developing experiments. These include deriving the most useful empirical hypotheses, thinking about independent and dependent variables, pretesting designs and procedures, and analyzing experimental data. Power analyses to decide needed sample size and experimenter effects on the data receive extended treatments.

Chapter 7, by Kathy J. Kuipers and Stuart J. Hysom, describes a number of practical problems that can occur in setting up experimental research and provides solutions for them. Challenges faced include everything from issues of experimental design to dealing with others in the organizational environment, what to do when scheduled participants fail to show up, and managing the funds to pay for their participation. This chapter includes discussions of ways to assess whether a participant was actually in the social situation that the experimenter wished to create or whether disbelief or misunderstanding seriously affected his or her interpretations and behavior. It also offers suggestions for ways to enlist participants in the research enterprise so that they do not, for instance, tell future participants details about the project that would skew the outcomes.

# Chapter 1

# Why Do Experiments?

**Murray Webster, Jr.**

*University of North Carolina—Charlotte, Charlotte, North Carolina*

**Jane Sell**

*Texas A&M University, College Station, Texas*

## I A BRIEF HISTORY OF EXPERIMENTS

Many social scientists, and most physicists, chemists, and biologists, see experimental methods as one of the defining characteristics of scientific inquiry. Although the experiment is far from being the only research method available to the social sciences, its usage has grown remarkably in the years since World War II. Many historical changes are associated with the growth in experimental methods, of which two kinds of changes are especially important: new topics and new technology.

In early decades of the 20th century, sociologists were largely occupied with classifying types and growth of societies or with development of different parts of cities. Following that war, many social scientists became interested in phenomena that can be studied experimentally. In sociology and social psychology, for instance, topics such as interpersonal influence, distortions of judgment, and conformity processes seemed more pressing than they had seemed before authoritarian and repressive societies were common topics. With the new topics came new theories, many of them amenable to experimentation. Economists began to conceptualize strategic game-playing and became interested in behavioral economics; political scientists developed rational choice theories of voting choice; communications scientists began to understand influence processes; sociologists had new theories of social exchanges; and psychologists, whose discipline had used experiments from its beginnings, expanded its study of effects of social factors that appear in the presence of one or more other individuals. New topics and new theory were both congenial to the development of experimental methods in social science.

The second factor was new technology. Starting at a few universities, experimental laboratories were built, followed by laboratories in government and at private research firms. New laboratories required and facilitated development

of many kinds of technological advance: coding schemes to record discussion groups, one-way mirrors and, later, television and computers to observe and control communication among researchers and experimental participants, sound and video recorders, and many other elements of contemporary experiments began developing in the years following World War II.

Although experiments are a recognized part of today's social science research techniques, for many social scientists they are still not well understood. Training in laboratory experimentation is still not part of the graduate training of the majority of social scientists (psychology may be the exception). That is unfortunate for many reasons. Those who might wish to conduct experimental research may not feel confident enough in their skills to approach this method. Social scientists who use other methods, such as survey researchers in sociology, may misunderstand the goals and uses of experiments. Because every science relies on peer review of research, misunderstandings can slow the accumulation of knowledge; good experiments may be criticized on inappropriate grounds, and real flaws in an experiment may be overlooked.

With the continuing growth and development of experimental methods in social science, it will not be long before understanding experiments is an important part of every social scientist's professional skills. We and the other authors in this book hope to contribute to that understanding. For new experimenters, we offer suggestions to improve the quality of their work; for those who read and wish to assess experimental research, we describe techniques and offer guidelines. In these ways, we hope to contribute to the growing quality of experimental research in social science. A poorly designed experiment will either produce no results or, worse, will produce results that are not what an experimenter thinks they are. This book brings together "best practices" by several of today's outstanding experimental researchers. The chapters can be read as "how to" manuals for developing one's own experiments or as sources of criteria to judge and improve the quality of experimental research by practitioners and by the professional audience. All of the chapters contain background to their individual topics that explicitly address common and some uncommon points crucial for understanding this method.

Experimental research is one kind of intellectual activity. A good way to approach experiments, either those one plans to conduct or those conducted by others, is to ask what they contribute to knowledge. What do we know as a result of an experiment or what do we hope to learn from a contemplated experiment? As will be clear in several of the chapters, the central issue in experimental research, as well as in other kinds of research, is how the research can contribute to knowledge of social processes and social structures.

We begin with some terminology on research design. All research is about how things are related. In describing a research design, often it is convenient to distinguish *independent* and *dependent variables*, and we use those terms in describing what we mean by *experiments*. A *variable* is anything that takes on different values, something that can be measured. In research, variables are tied

to measurement operations; for instance, the variable *socioeconomic status* or SES may be measured by a person's or a family's income in dollars. Beginning students must become accustomed to the idea that so-called *independent variables* are usually controlled in some way by investigators, whereas *dependent variables* are left free to vary; they are controlled only by nature. Thus, a study might control educational level statistically by partitioning a sample of individuals into those whose education ended after eighth grade, after some high school, with a high school diploma, etc., to determine education's effect on SES. Education is the independent variable in this study, and SES is the dependent variable. The design described uses survey methods.

As we use the term, a study is an *experiment* only when a particular ordering occurs: when an investigator controls the level of an independent variable(s) *before* measuring the level of a dependent variable(s). In the preceding hypothetical survey design, presumably data were collected all at the same time on the respondent's schooling and income. The independent variable was partitioned afterwards, and the interest was in how SES divided after education was so divided. If the design had been an experiment, the investigator might have begun with a large group of children and placed them into different groups determining how many years of education they would receive. When members of the last group had completed their education, average income levels of the different groups might be compared.

That hypothetical example illustrates two more points about research designs. First, not every research design is an experiment. Informal usage sometimes describes any study as an experiment, but if we are going to focus on creating good research designs, we need to be clear what type of designs we work on. Although many of the criteria of good scientific research apply to all kinds of research, there are specific criteria for good surveys that are quite different from the criteria of good experiments. For instance, sample representativeness is a major concern in many surveys and less so (for reasons discussed in several chapters) for experiments. Random allocation of respondents to conditions is crucial for most experimental designs, but it is often impossible in surveys. Second, for many reasons, not every research question can be studied experimentally. Moral and practical considerations make the preceding hypothetical experimental design—deciding ahead of time how much education every person can attain—ridiculous. Knowing what we can study experimentally is important at the very earliest stages of research design, when an investigator selects a type of design to develop.

Experimental studies came into social sciences in approximately 1900, beginning in psychology with studies of biological responses (particularly saliva production in dogs) conducted by the Russian physiologist Ivan Petrovich Pavlov. Pavlov's training in the biological laboratory was demonstrated in his turning to an experimental design, as well as in his use of physiological terms such as "stimuli" to name independent variables in the work. American psychologists such as Edward L. Thorndike (1905) at Harvard University and

J. B. Watson (1913) at Johns Hopkins University adopted and developed the method for many studies of individual differences and interpersonal influences, and the methods spread across the social sciences.

Social psychologists Solomon Asch (1951), Muzafer Sherif (1948), and Leon Festinger and J. Merrill Carlsmith (1959) developed experimental methods beginning in the 1940s, approximately the same time the economist E. H. Chamberlin (1948) began to study markets experimentally. Mathematically trained social scientists, including Siegel and Fouraker (1960) and Von Neumann and Morgenstern (1944), analyzed rational choices, negotiations, and games, providing the foundations for many contemporary theories in sociology, political science, communications, and economics. In the 1950s, Robert Freed Bales and his colleagues and students at Harvard University began studying discussion groups using techniques and technology of an experimental laboratory. Kurt Lewin and Dorwin Cartwright founded the Research Center for Group Dynamics at Massachusetts Institute of Technology in 1945, and Cartwright moved it to the University of Michigan upon Lewin's death in 1947 (Cartwright & Zander, 1953). The experiment has been a significant part of all the social sciences for more than half a century, and its particular advantages continue to attract researchers across the social sciences.

The history of experimental methods shows diffusion across disciplines. Starting with techniques Pavlov learned in biology, through widespread diffusion of Bales' studies of discussion groups (e.g., into contemporary "focus groups" in communications and business), imaginative researchers have built on what worked in other projects and other disciplines. This is all to the good. When investigators in one discipline come up with useful designs, other disciplines benefit from using those when appropriate. Designs and technology do not have to be reinvented constantly, and investigators can focus their attention more appropriately on research questions. Experimental methodologies are not tied to any particular discipline; it makes no sense to argue, for instance, that economic experiments are fundamentally different from experiments in psychology or sociology.

There are different theoretical concerns in different disciplines, and different disciplines have developed different typical designs and even display different "tastes" in design. For instance, some experiments in social psychology use deception to create independent variables, something most economists do not do and often do not approve of. Importantly, however, there definitely are criteria for well-designed experiments and reliable experimental results, and those do not differ across disciplines. Quality experimentation is furthered by ecumenism.

## II COMPARISONS TO OTHER DESIGNS USED IN SOCIAL SCIENCE RESEARCH

Social scientists have a wide range of data collection methods. They include the following:

- Unstructured observation. This method is typically used when a social scientist is present at some unexpected event, such as an accident, natural

disaster, or violent attack. The scientist can record relevant data on presumed causal factors and outcomes, often also identifying intervening factors and their effects. Other observers, such as from news media or ordinary citizens, can also record unstructured observations, although in many cases the social scientist's training helps him or her to choose what to observe and record as being theoretically important.

- Structured observation. This method is used when a social scientist begins with a coding scheme of what to observe, perhaps also how to record it. In contrast to unstructured observation, here the observation site is predictable and well enough known so that observations can be chosen ahead of time. For instance, a sociologist might go to a courtroom to code jurors' facial expressions and other behaviors to determine how those may be related to verdicts. The observations here are limited and focused, and it is possible to assess reliability if two observers record the same data from the situation.
- Historical archival research. This method relies on documents as data for answering research questions. For instance, a researcher may compare recorded lynchings to the price of cotton in the old Confederacy states. Documents may be written, video, or audio. As with observational studies, archival documentary research does not exert control over factors, except through statistical techniques.
- Participant observation. This may be structured or unstructured; its defining characteristic is that the observer has privileged access to the setting by virtue of being a member of the group studied. The natural settings are extremely important for such studies; their explication is as high a priority as the interaction itself. So, for example, the rules of the sheltered workshop, the structure of the asylum, or the city in which a gang operates are important actors in their own right. In many ways, observation research of all three types is at the other end of a methodological continuum from experiments: the setting is natural, and the outgoing processes in actual settings are valued. Control and randomization, so highly prized in experimental research, are not usually possible in such studies.<sup>1</sup>
- Survey research. Surveys are usually defined as observations on or about individuals or groups. The U.S. Census is a famous survey used by literally millions of researchers. It asks respondents about many behaviors or characteristics, and then analysis proceeds by controlling some factors to determine how that affects outcome variables. Survey research is generally interested in generalizing from a sample to a particular empirical population. Interestingly, an incorrect presumption that generalization proceeds from experiments and from survey samples in the same way can lead to survey researchers criticizing experiments for being "artificial." We consider

---

1. There are exceptions. For example, Milton Rokeach's (1981) *The Three Christs of Ypsilanti* involved a kind of control because all three men claiming to be Christ were in the same setting and talking to the same psychologist (Rokeach).

- artificiality in greater detail later, as well as the differences between sample generalization in surveys and theoretical generalization from experiments.
- Combining different methods is a defining characteristic of case studies. Case studies often involve documentary historical data and may also include surveys and participant observation.<sup>2</sup>

### III ADVANTAGES AND DISADVANTAGES OF EXPERIMENTS

The benefits of any research method cannot be assessed independently of the questions the method is designed to answer. A beautiful research design cannot compensate for a flawed research question. This is especially true for experiments because they are designed to determine how specific kinds of independent variables and antecedent conditions affect dependent variables (or consequents). If a research hypothesis, for example, is derived from faulty assumptions or premises, even the most elegant research—experimental or otherwise—cannot save the study. What a good design can do, however, is help the investigator identify the part of the theory that is faulty. Disconfirmed predictions, when they have been derived from an explicit theory, are valuable because they can show which assumptions or conditions in the theory need improvement.

The greatest benefits of experiments reside in the fact that they are artificial. That is, experiments allow observation in a situation that has been designed and created by investigators rather than one that occurs in nature. Artificiality means that a well-designed experiment can incorporate all the theoretically presumed causes of certain phenomena while eliminating or minimizing factors that have not been theoretically identified as causal. In the language of research design, an experiment offers an opportunity to include the independent variables of theoretical interest while excluding irrelevant or confounding variables. An experiment (like a theory) is always simpler than a natural setting; because of that, it offers the possibility to incorporate factors of interest and limit extraneous factors so that the theoretical principle being examined is isolated. Experiments permit direct comparisons: most often the comparison is between conditions in which a factor is present (an experimental condition) or in which the factor is absent (a baseline condition).<sup>3</sup> In this way, the effects of a factor in the experimental condition can be gauged. Such direct testing of a theoretical prediction is much more difficult in other, less artificial settings. Because they

---

2. Robert K. Yin (1984) argues that case studies can share some of the characteristics of experiments. In particular, he argues that multiple case studies can be chosen such that they represent variation on the independent variable. Then measurements can be taken on the dependent variables. So, for example, if the researcher had a theory concerning how land use compared between government-mandated programs and voluntary programs, she could randomly choose (or perhaps choose based on specific criteria) mandated and voluntary programs, gather data, and examine the dependent variables.

3. Sometimes comparison may be between an empirical condition and a theoretical prediction rather than between two empirical conditions.

are less artificial, they are more complicated and contain a myriad of factors that could interfere with, magnify, or dilute the effects of the particular factor being investigated.

Another technique that experiments often use to ensure comparison is random assignment. The power of randomization is the power assured by probability theory: if extraneous influences (errors) are distributed randomly, they sum to zero. When individuals are randomly assigned to different experimental conditions, different effects observed in different conditions are not due to uncontrolled factors, such as personal traits of the individuals studied, because those factors have been evenly distributed across conditions. So, for example, in drug studies, there is often a control condition or treatment in which individuals receive only a sugar pill and an experimental treatment in which individuals receive only a treatment drug. When individuals have been randomly assigned to either the control or the treatment condition of the experiment, comparing outcomes in the control or baseline (sugar pill) condition with the treatment (drug) condition permits a gauge of the effect of the drug treatment. Random assignment of individuals to conditions assures a researcher that there is nothing unanticipated about the individuals that might have led to an effect. So, for example, if there is no random sampling, there is the possibility that, by some chance, healthier people or people with better nutrition or taller people ended up in the same condition. With random assignment, we can eliminate the explanation that differences in conditions result from characteristics of the individuals or group treated.

Artificiality also allows experiments to provide settings that are difficult, if not impossible, to find. For instance, some studies of voting, such as the experiment of Wilson (see Chapter 14), allow individuals to record dozens of votes within a single experimental session to assess a theory of how individuals reach equilibrium points where each person's interests are represented as closely as possible. The situation is analogous to repeated voting on bills in Congress or to repeated voting among condominium residents. However, because Wilson studied it experimentally, he was able to assess the process not in months or years but within a couple of hours per group.

Because experiments are artificial and controlled, they also invite and enable clear replication by other investigators and comparison across different settings. Findings from an experiment can be assessed by someone else who replicates the experiment. In contrast, findings that occurred in some natural setting can be impossible to replicate because it can be very difficult or impossible to find another natural setting enough like the first one that the investigator can be confident it truly replicates the important features of the setting. Also, experiments are well suited to cumulative research programs that develop and test theories sequentially because they can be evaluated under consistent conditions. The researcher can be assured that differences do not arise over different settings or different operationalizations but, rather, because of the different theoretical factors being tested. Such consistency can be crucial for theoretical cumulation—results for one study can be used for subsequent studies. For instance, the

experimental or manipulated condition in a study at time one can become the baseline condition for study two.

Temporal or theoretical ordering also can be examined in experimental settings, brought by, again, artificiality. In interactions that occur in natural settings, it is often difficult to disentangle cause and effect, antecedent and consequent conditions. Thus, for example, it is difficult to disentangle how experience within a group and the statuses of individuals within a group (often derived from their experience) affect individuals' behavior. In experiments, not only can experience and status be separated but also they can be manipulated prior to the measurement of behavior. In this manner, the antecedent and the consequent can clearly be distinguished. In natural settings, this is much more difficult, even given longitudinal designs, because individuals possess many different characteristics and experiences, those factors may change dramatically over time, groups may be involved in different kinds of tasks, and group composition is constantly changing.

Although enormous benefits accrue from artificiality, experiments receive a great deal of criticism for this defining property (e.g., see [Babbie, 1989](#)). The criticisms center around generalizability: because experiments are artificial, they do not mirror any real setting and they are not representative of a particular empirical population. Those criticisms are correct, as far as they go. Furthermore, they might be considered disadvantages of experiments. We argue that is true only insofar as experiments are not appropriate for certain kinds of questions. Basically, if the goal is to study properties of some natural setting itself, an experiment is not particularly appropriate. For studying theories that have abstracted some properties of natural settings, however, experiments can be ideally suited.

Experiments cannot attempt to simulate all the complexities of particular settings. They cannot reproduce, for example, elementary school classrooms in Texas or all features of the work environment of a particular corporation such as Apple. However, experiments can produce abstract features of school classrooms (e.g., giving someone chances to perform and then evaluating performances), and they can produce authority structures such as might also occur in a corporation such as Apple. So, rather than reproduction, experiments are designed to match the characteristics of theories composed of precisely defined, abstract concepts. In Chapter 8, Zelditch develops this argument in detail in his analysis of external validity.

These concepts and, consequently, the theories composed of these concepts are not defined by a particular time or place. Rather, they are abstracted from particular times and places. For instance, social scientists from several disciplines have studied "public goods," a line of research described by Sell and Reese in Chapter 10. Naturally occurring instances of public goods might include parks in a large city, listener-supported radio, or a free shuttle bus. Those instances have certain features in common. Individuals need not pay money in order to enjoy their benefits—anyone may be allowed to use a park or a playground in

the city. However, there always are costs, such as taxes foregone by keeping businesses out of the park, and cost of maintenance crews. Public goods dilemmas revolve around keeping enough people motivated to contribute a fair share for the things they can actually use without paying.

The term “public good” is an abstract concept meant to capture key features of concrete things such as buses and parks. The term is independent of time and place: whereas a shuttle bus may be an instance of a public good at the present, the term public good will apply in any culture at any time where other concrete things may meet that definition. A theory using the term public good must include an abstract definition telling someone how to find instances of it. The definitions of abstract terms can be much more precise than definitions of concrete terms. Although someone might argue whether a particular piece of land is or is not a public park, for instance, if the theorist has done his or her job well, there will be no doubt whether something is a public good.<sup>4</sup> In Chapter 3, Thye provides more detail on abstract definitions and the parts of a theory. Our point here is that experiments are ideally suited to developing and testing abstract theories, and this is their most appropriate and most valuable use. If there is no good theory, it is too early to think about doing experimental research. The investigator would be better advised to concentrate on observation of natural settings for a while.

Because the theories are abstract and the experiments are artificial, the question arises: how can the results and general principles be applied to the settings in which we really are interested—natural settings such as those elementary school classrooms in Texas? The answer rests with the chain of activities in scientific research and application. At the first stage, an investigator decides on a particular kind of research issue, problem, or theoretical dilemma. The issue could arise from observations of some phenomenon that, for whatever reason, she wishes to understand better.

However, observation is not a necessary prerequisite. The research issue could be suggested by elaborations of other theories or by purely deductive implications of a formal system. The investigator conceptualizes the phenomenon abstractly to develop propositions about how it functions under different conditions. In other words, she comes to develop a theory. Next, she may create an experiment to test her developing theory in exactly the kind of situation that illuminates the parts she is most concerned about. If the experiment receives experimental confirmation, she comes to believe in her developing theory and is ready to apply it to understand other natural settings. She might even, if she is unusually energetic and skillful, decide to use her theory to devise interventions to produce desirable changes in some natural setting that fits the relevant conditions specified in the theory.

---

4. Logicians say such definitions are “exact class,” and they obey the law of the excluded middle. That means that for an adequate definition, it must be clear whether something does or does not meet that definition; there is no middle ground where it “sort of” meets the definition.

Note that the theory is the bridge between the laboratory experiment and natural settings. The experiment tests the theory, and the theory can be applied to natural settings as well as to the experiment.

Thus, for example, suppose that a theory predicted that norms for cooperation in dyads would strongly increase cooperation of a whole group (composed of many dyads) in a public good context. Further suppose that a scope condition for this theory was that the initial dyadic cooperation occurs in settings in which there are no authority figures endorsing the cooperation. Such a theory could be applied to classroom settings as long as no individual (e.g., a teacher) required dyadic cooperation.

## IV STEPS IN CONDUCTING EXPERIMENTAL RESEARCH

Experimental work takes place in four large blocks that we call *foundations*, *abstract design*, *operations*, and *interpretations*. *Foundations* include theory and hypotheses behind an experiment; they are the intellectual reasons to conduct the experiment. General questions at this stage are “What do I want to learn and how can an experiment answer those questions?” *Abstract design* refers to the “plan” of an experiment, including independent and dependent variables, measures, interaction conditions, and all other things that go into making up an experiment. The main question here is “How can I design a situation to answer the previous research questions?” *Operations* are the way things actually look to someone watching or participating in an experiment. Variables become operational measures; interaction conditions become instructions or words used by participants to communicate with each other. An important question to consider here is “How will this look in actual experience to someone in the experiment? Is that the same as I intended in my design?” *Interpretation* is what experimental results mean. The general question at this stage is “What did we learn from the experiment?” The answer may be in terms of the theory or hypotheses stated at the foundations stage or in terms of some unrelated general issues such as methodological advance.

All four kinds of activity—and all four questions—must occur before an experiment is actually conducted. Although that might be obvious for the first three blocks (*foundations*, *design*, and *operations*), it is equally true for *interpretations*. That is, an experimenter should consider what results of an experiment could mean before beginning the work. Experiments that work out—that is, that support research hypotheses—have fairly straightforward interpretations, but experiments resulting in disconfirmed hypotheses also mean something. We expand this point later. Let us consider in greater detail what each of the four kinds of activities entails.

*Foundations* of an experiment are the reasons for doing it. An experimenter has some questions to answer or a set of ideas to assess. Ideally, the experimenter begins with a developed theory and rigorously derived hypotheses, although often theoretical and hypothesis development are incomplete. However, by understanding the ideal case, often it is possible for an investigator

to assess how damaging departures from that ideal may be—that is, whether the work is ready for experimental investigation or whether more time needs to be spent on the theoretical foundations of the experiment.

Theoretical foundations have different elements that affect experimental design. We consider scope and initial conditions and derived consequences or hypotheses. (Other essential parts are the abstract propositions that constitute the actual theory and the logical or mathematical calculus used to combine propositions to yield derivations or hypotheses. However, scope and initial conditions and hypotheses are crucially linked to experimental design, and we focus on those parts here.)

All theories have limited scope. Scope conditions describe classes of situations to which a particular theory claims to apply. Newton's laws of acceleration of falling bodies (the general propositions) famously apply only in the absence of factors such as air resistance or magnetic deflection; in other words, they take as scope conditions the absence of those factors. Einstein's special theory of relativity treats the speed of light as a constant; it does not use scope limitations of air resistance or deflection. Sociological theories of status processes treated in Chapter 12 take as scope conditions situations in which individuals are task-focused and collectively oriented—that is, situations in which individuals view solving problems as their main reason for interacting (rather than, for example, enjoying each other's company) and in which they believe it is important to let everyone contribute to problem solving (as opposed to taking tests without help).

Scope conditions tell the kinds of situations in which a theory claims it can describe what happens. If a situation meets its scope conditions, the theory should be able to predict accurately. If a situation is outside that scope, the theory makes no claim to being able to make predictions. Given that scope conditions are met, confirmation of prediction increases confidence in the theory and disconfirmation indicates something wrong with the theory (or the test, if methodology could be a problem). Thus, confirmation and disconfirmation both have meaning *as long as a situation is within the theory's scope*. Outside the theory's scope, any results are irrelevant for judging the theory. This means that an experimenter must be careful to design and operationalize situations within a theory's scope.

It is unfortunately true that many social science theories have been offered without any explicit scope conditions. They are presented as if they applied to every conceivable kind of social situation. If you pressed a theorist about that, he or she might admit that the theory probably does not apply everywhere, but many theorists appear not to have considered just where their theories do and do not apply. To design an experiment to assess a theory that is not explicitly scope-limited, an experimenter must infer (i.e., guess) what kinds of scope limits the theorist had in mind. In other words, suppose a particular natural situation or experiment produced results different from what the theory would expect. Would the theorist say, “Well of course I never intended the theory to apply to such cases?” If so, that is an implicit scope condition. All theories have scope

limitations. Although it is really a theorist's job to tell what those are, if a theorist does not do that, the experimenter must because there is no way to design a relevant experiment until issues of scope have been settled.

Initial conditions are what prompt the theoretical process; they are the instigators of change in a situation. Presuming a situation meets a theory's scope conditions, then initial conditions describe the setup for what will happen. For example, in theories of status processes mentioned previously, one situation of interest is what happens when two individuals occupying different status positions interact. Interactants having different statuses (however their society may define status) is an initial condition in this case. In that theory, if individuals have different status positions, that fact will become salient to their interaction, they will form differential performance expectations for each other, and they will treat each other differently in specific ways, all related to power and prestige inequality. From the theory, we could derive predictions or hypotheses that if a high-status person interacts with a low-status person, the former person will talk more and listen less, and the latter's discussion behavior will be the inverse. The initial condition to be created in an experimental test is status inequality.

Initial conditions often define the independent variables of experimental design. The preceding predictions might be tested in an experiment designed so that two individuals of different status positions will interact in a task-focused collectively oriented situation. To be clear, the theory states that if a status difference exists, then certain behavioral outcomes will occur, given a situation meeting the scope conditions. An experiment to assess that prediction must, therefore, create a situation in which individuals of different statuses encounter each other. The design is governed by initial conditions specified in the theory and by its predictions. Independent variables must reflect initial conditions, and dependent variables that reflect predicted theoretical outcomes must be possible.

*Experimental operations* are what actually will happen in an experiment. They are the instances an experiment offers of the abstractly defined independent and dependent variables. In the previous example, creating a situation in which two individuals differing in status interact might be accomplished by creating a team consisting of a teenager and a 35-year-old woman. In order to use that operation, we would need to know that age constitutes a status characteristic for people in our society because the theory and the independent variable speak of "status difference," not "age difference." The theory predicts effects of status processes; it is definitely not a theory of age (or of gender or skin color or any other actual characteristic). The theory states that *if* age is a status characteristic for people in a particular society, then it will have certain interaction effects. The operational challenge is to instantiate status difference, which in our society can be done using age difference.

For a theorist, the distinction of the theoretical term "status" and the operational term "age" is important because societies change. Status characteristics are created and atrophy all the time, but the same theoretical processes apply to whatever may be the status distinctions existing at a particular time. The

theoretical status process, not the historical fact of what happens to be status, is of primary interest here.

For an experimenter, the distinction of theoretical terms and operational terms is also crucial because a well-designed experiment must instantiate “status difference” in order to test that theory. In other words, many challenges at the stage of operational design involve creating situations that are instances of cases described theoretically. Partly that means understanding the culture in which the experimenter operates—for instance, what constitutes a status characteristic for this population.

However, operational challenges extend far beyond what an experimenter knows of society. He or she must also translate crucial concepts into terms meaningful for the experimental participants. For instance, if a situation must be task-focused, as one of the scope conditions requires, that must be created in any experimental test of that theory. Yet experimental participants do not ordinarily use a term like “task-focused,” and it would not communicate much to tell them to try to simulate being task-focused. The experimenter must create a situation that individuals will accept as task-focused. In other words, experimenters must understand the world of participants well enough to instantiate situations they understand in the same ways as the experimenter does, even though the experimenter thinks in abstract, theoretical terms and participants think in concrete actual realities. Going back and forth between the theoretical world and the worlds of participants is one of the most difficult challenges facing experimenters.

Operations include introductions and instructions—whatever participants are told when they appear for an experiment. Of course, those elements should be as standard as possible because in a well-designed experiment, participants in different conditions experience most of the same situation, with the exception of differences specified by independent variables and initial conditions. Reading instructions, for instance, is preferable to talking about a situation informally because the former kind of administration is much more uniform.

The final part of operations is dependent measures—operational translations of dependent variables from the guiding hypotheses. Here again, translation is not straightforward. An experimenter has to think about what can be done technically and given the limits of human observers. Limits include technology, such as what can be seen, and human limits such as fatigue and inattention. In our status example, the hypothesis predicted more talking and less listening from high status people. How can “more talking” be assessed? By recording their interaction and counting words? Maybe sentences would be a better measure, or maybe complete thoughts. How can an experimenter measure amount of listening? Is silence enough, or does the hypothesis entail some sort of paying attention, and if the latter, how can that be measured?

None of the steps in creating operational measures is likely to be simple, and yet the strength of any experimental design depends on doing all of them well. The general advice here is for the experimenter to think through absolutely

every step in the operations as thoroughly as possible and then pretest everything to be sure it works as intended.

A crucial decision in experimental operations involves what data will be collected and how they will be used. In an experiment, it is possible to assess behavior of various sorts, to distribute questionnaires, and to interview participants. Selecting measures to use and carefully developing them are significant steps in experimental design. Just the simple case of testing a prediction about who will talk more, for instance, requires making decisions about how much difference matters. If an experimenter counted words, would a 10% difference in average words across high-status and low-status conditions, for instance, be considered confirmation? Could we put the decision in the hands of statisticians so that any statistically significant difference counts as success? In addition to mean differences across conditions, many other statistics are available that have potential value. For instance, variability across participants may give clues to both theoretical and methodological questions.

Presuming an experimenter has reached thoughtful decisions about what sorts of data to collect and what will constitute confirmation and disconfirmation of hypotheses, other questions arise. Usually, there are many hypotheses possible from a theory under test, and in the nature of things, not all of the test results will point in the same direction. That is, it is reasonable to expect some confirmations and some disconfirmations. An experimenter who gives some thought to what different possible patterns might mean before collecting the data is in a better position to collect additional information, such as other measures, when disconfirmations or indeterminate outcomes appear.

Even the purest cases—total confirmation and total disconfirmation—deserve some advance thought. Total confirmation of all hypotheses greatly strengthens confidence in the theoretical foundation, of course. This happy state should lead to the question “What’s next?” An experimenter who has even a sketch of an idea what questions to investigate following the present experiment is primed to move quickly once the data are in.

What is not often appreciated is that disconfirmation is also informative. It might, of course, mean an entire theory, or at least the parts of it under test, was wrong. However, that outcome is superficial and it is never likely. Did the experimenter really begin with a theory so weak that he or she is not surprised by getting total disconfirmation? That seems unlikely. Most often, experimenters believe the theories they work with, and disconfirmation does not usually mean a theory should be abandoned. Abandoning a theory without a substitute theory really says “I have no idea at all what’s going on here.” Although that might be true in rare instances, for most social situations, we have at least vague intuitions about what is going to happen. (The work of an experimenter, as of a theorist, is to make vague intuitions explicit and testable.) Disconfirmation invites further thought. What parts of the theory were used in the disconfirmed prediction? How could those parts be modified to account for the outcomes of the experiment? What would an independent test of that reformulation look

like? In a well-designed experiment, disconfirmation is at least as useful information as confirmation.

Much of this section identifies challenges and potential problems with good experimental design. That reflects the world: good experiments are difficult to design and conduct, and only a foolish person thinks otherwise. Fortunately, some of the best experimenters in social science have prepared guidance for all steps of the process in the succeeding chapters of this book.

## V THE PLACE OF EXPERIMENTS IN SOCIAL SCIENCE

Experimental methods are one kind of data collection that may be used to assess theoretical knowledge. They are certainly not the only method, for many others are available: surveys, content analyses, structured and unstructured observation, and others. As noted previously, all research searches for relations between concepts. All good research depends on being able reliably to infer that things are related in the ways a theory thought they would be—or in understanding what parts of the predictive apparatus needs revision.

The preceding discussion suggests that experiments are not well suited to study in the absence of any theoretical foundation—for instance, just “to see what will happen.” Anytime someone sets up a situation and collects observations, we know that “something will happen.” However, the real questions are what we learned from it. Did something happen for a reason we can specify, or was it just a chance occurrence? If the latter, it is not interesting scientifically. If there is a reason, that is the beginning of a theoretical understanding. A scientist will usually want to work out the theoretical understanding before moving to empirical research. The reasons why are more interesting and important than the simple facts of what happened.

In fact, a reader may have realized another feature of experimentation from the discussions in this chapter: experimental results themselves are not really interesting except as they bear on a theory. In other words, who cares if two people in our status example talk different amounts? That becomes an interesting fact if it shows us something about how status operates, for we will encounter thousands of other situations of status inequality and with the theory, we have some way to understand what is likely to occur in those new situations also. However, the simple numbers from experimental data hold little interest without a theory. (For other kinds of research, such as unstructured observations of police interacting with drug dealers, the observations themselves are likely to be interesting; the cases may be described as “socially important.”) For experimental research, the theory is the interesting part; the experimental results are interesting insofar as they tell us something about the theory.

Currently, some phenomena do not lend themselves well to experimental research. Understanding how different patterns of parent-child interaction affect children’s success in school, for instance, may require several years of observation in natural settings. However, sometimes the apparent inapplicability of experimental methods becomes less serious upon closer examination. Many

years ago, Morris Zelditch, Jr. (one of the authors of this volume) wrote an article titled, “Can you really study an army in the laboratory?” (Zelditch, 1969) and answered in the affirmative. Although you cannot take an army into a laboratory, you certainly can study important theoretical features of armies, such as authority structures, power exercise, and organizational efficiency.

As the social science become more theoretical, experimental methods are likely to become increasingly important. Experiments generally offer the most convincing evidence for success or weakness of theoretical explanations. No other kind of research method produces data so directly relevant to a theory or suggests causality so conclusively as do experiments.

## VI HOW THIS BOOK CAN HELP

Most students first learn about experimental methods from coursework in colleges, but there is a long way to go from knowing experiments exist to being able to actually do one and get some useful results. It is unreasonable to expect a student who has studied chemistry to be able to go into a laboratory and create a drug to block the replication of HIV or build a safe cleaning solution for home use. Nobody would just give a biology student a preserved animal and ask her to figure out why it died. It is the same in the social sciences. There are some theoretical ideas about how social structures and social processes work, but those theories do not tell enough for someone to be able to study those things experimentally.

Laboratory experimentation is taught in all fields through apprenticeship programs. Students, from new undergraduates to postdoctoral trainees, work in laboratories with established scholars. At the early level, students learn some very basic facts about experimental methods, such as Mills’ canons of difference and similarity, the importance of developing theoretical concepts and tests, and how to use and interpret statistical tests. At the most advanced level, a student learns advanced techniques, how to use and interpret very sensitive instrumentation, an extensive body of experiential knowledge about likely outcomes and what they mean, and other esoteric knowledge usually associated with a narrow specialization.

The odd fact is that for the vast area between the extremes of expertise, there is little in the way of reliable information. This means that someone who has never designed an experiment is unlikely to find a good source of the information needed to get useful outcomes from it. Even an experimenter in one specialty is unlikely to know how to devise and conduct useful experiments in another area. Because social science experimentation is newer and less widely known than natural science experimentation, many of us do not have a good understanding of why someone would want to do experiments, what they are good for, what problems can arise and how to deal with them, and what some exemplars of good experimentation look like in the different social sciences. This is the gap we address in this book.

## REFERENCES

- Asch, S. (1951). Effects of group pressure upon the modification and distortion of judgment. In H. Guetzkow (Ed.), *Groups, leadership, and men* (pp. 177–190). Pittsburgh, PA: Carnegie Press.
- Babbie, E. (1989). *The practice of social research*. Belmont, CA: Wadsworth.
- Cartwright, D., & Zander, A. F. (1953). *Group dynamics: Research and theory*. New York: Harper & Row.
- Chamberlin, E. H. (1948). An experimental imperfect market. *Journal of Political Economy*, 56, 95–108.
- Festinger, L., & Carlsmith, J. M. (1959). Cognitive consequences of forced compliance. *Journal of Abnormal and Social Psychology*, 58, 203–210.
- Rokeach, M. (1981). *The three Christs of Ypsilanti: A psychological study*. New York: Columbia University Press.
- Sherif, M. (1948). *An outline of social psychology*. New York: Harper & Brothers.
- Siegel, S., & Fouraker, L. E. (1960). *Bargaining and group decision making*. New York: McGraw-Hill.
- Thorndike, E. L. (1905). *The elements of psychology*. New York: Seiler.
- Von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton, NJ: Princeton University Press.
- Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychological Review*, 20, 158–177.
- Yin, R. K. (1984). *Case study research: Design and methods*. Newbury Park, CA: Sage.
- Zelditch, M., Jr. (1969). Can you really study an army in the laboratory? In A. Etzioni (Ed.), *Complex organizations* (pp. 528–539) (2nd ed.). New York: Holt, Rinehart, & Winston.

## Chapter 2

# Ethics and Experiments

Karen A. Hegtvedt

*Emory University, Atlanta, Georgia*

## I INTRODUCTION

The role of ethics in research is no longer simply a matter of concern for individual researchers but, rather, a matter of public discourse. Regulations governing the protection of the rights and welfare of human research participants shape that discourse. The necessity of regulatory compliance, however, does not diminish the importance of underlying ethical concerns. This chapter contextualizes federal requirements in terms of ethical issues and principles. Discussion focuses on four general issues regarding the ethics of research in the social sciences (objectification, potential harms, exploitative practices, and confidentiality) and two issues specific to experiments (deception and subject pools). Ways in which researchers may deal with these issues in their attempts to meet federal regulatory requirements are outlined. Increased awareness of ethical issues and knowledge of regulations ensure the safety and well-being of individuals who participate in experiments as well as maintain the public's trust in the scientific endeavor.

Social scientists investigate a variety of social processes (e.g., bargaining, power, status, justice, trust, cooperation, conflict, attitudes, and decision making) using laboratory experiments. Although some may contend that such methods fail to capture the “real world,” as noted elsewhere (Martin & Sell, 1979; Webster & Sell, Chapter 1, this volume; Zelditch, 1969), it is not the intent of experimenters to recreate the real world in the laboratory. Undoubtedly, however, many elements of the real world affect the development and execution of laboratory experiments. One very important element—beyond the vagaries of funding and the challenges of operationalizations—is ethics.

The post-World War II era saw the transformation of ethics in research from a matter of concern for individual researchers to a matter of public discourse (McBurney & White, 2012). The emergence of modern ethical codes stems from the standards resulting from the Nuremberg trial of Nazi doctors who conducted cruel medical experiments on concentration camp prisoners

(see Dunn & Chadwick, 2004). The Nuremberg Code introduces the requirement of informed consent for nontherapeutic research. Subsequent documents (e.g., the 1964 Declaration of Helsinki by the World Medical Association) and U.S. federal regulations (e.g., the 1974 National Research Act and subsequent policy, codified in 45 CFR § 46) extend the premises and safeguards of the Nuremberg Code. Concern with ethics in (especially biomedical) research is growing in many other nations, but it is not yet as formalized as in the United States. In addition, by the turn of the 21st century, nearly all U.S. professional associations within the social sciences had established their own codes of ethics.<sup>1</sup> Such codes represent contemporary consensus on what a profession believes to be acceptable practices. The associations, however, have no means of routinely regulating the behavior of individual researchers (Rosnow, 1997). In the United States, regulation falls under the rubric of institutional review boards (IRBs), established by universities and other organizations as specified by federal regulations.<sup>2</sup>

Although critics of IRBs may claim that guidelines for the protection of human research participants are more suitable to medical research than to social and behavioral studies (Citro, Ilgen, & Marrett, 2003; DeVries, DeBruin, & Goodgame, 2004; Israel & Hay, 2006), the history of social science research includes landmark studies that highlight the harm that can occur in both field and laboratory research on processes such as conformity and influence. Most frequently cited is Milgram's (1974) experimental study of obedience to authority in which adult volunteers in the role of "teacher" were led to believe that they administered electrical shocks when another ostensible volunteer in the role of "learner" failed to give a correct answer. Such intentional deception has been harshly criticized (see Baumrind, 1985). Ethical concerns, however, emerge with regard to studies even in the absence of deception. Zimbardo's (1973; Zimbardo, Banks, Haney, & Jaffe, 1973) examination of the behavior of randomly assigned volunteers to the roles of "prisoner" and "guard" in a mock prison setting who endured psychological and physical abuse or became sadistic, respectively, exemplifies such concerns. The debriefing that Milgram provided and Zimbardo's early halt to the mock prison study indicate researchers' recognition of the potential and actual harms stimulated in their studies. There was, however, little subsequent attention to the lingering effects of these harms.

---

1. Professional associations' ethics codes include the following: American Psychological Association (2010), <http://www.apa.org/ethics/homepage.html>; American Sociological Association (1999), <http://www.asanet.org/about/ethics.cfm>; American Political Science Association (2008), <http://www.apsanet.org/imgtest/ethicsguideweb.pdf>; and American Anthropological Association (2012), <http://www.aaanet.org/profdev/ethics>. The American Economic Association appears to have no explicit ethics code.

2. Federal Policy for the Protection of Human Subjects, 45 CFR § 46 (2001) may be found at <http://www.hhs.gov/ohrp/humansubjects/commonrule>. The OHRP (Office of Human Research Protections) also posts guidance for IRBs and researchers on a variety of issues (e.g., <http://www.hhs.gov/ohrp/policy/index.html>).

Thus, concern for welfare of study participants permeates social science as well as biomedical research. This chapter first addresses the meanings of ethics in aspects of research. Attention then shifts to general ethical issues in research and specific issues characteristic of laboratory experiments. The chapter also describes how IRBs operate with regard to the review of social and behavioral research and strategies for experimenters to meet federal regulatory requirements regarding protecting the rights and welfare of human study participants. Insofar as such protection captures the essence of ethical research, the strategies generalize to meet the requirements of other countries' ethics committees. Conclusions focus on the mesh between researcher aims and values and the demands of the broader context to ensure ethical research.

## II DEFINING ETHICS IN RESEARCH

Ethics broadly refer to the moral principles governing behavior—the rules that dictate what is right or wrong. Yet as the previously noted examples illustrate, what constitutes right or wrong is subjective, defined by groups with particular aims. Such aims underlie the fundamental conflict between (social) scientists who pursue knowledge that they hope may benefit society and the rights of research participants ([McBurney & White, 2012](#); [Neuman, 2011](#)). In the absence of moral absolutes, professional associations and others craft rules for what is proper and improper regarding scientific inquiry to ameliorate this conflict. The resulting ethics codes reflect philosophical ideas and attempt to bridge to regulatory requirements. The ethical conduct of research pertains to more than data collection involving human participants and encompasses more than simply complying with specific federal regulations protecting such participants.

Scientific misconduct discussions (e.g., [Altman & Heron, 1997](#); [Neuman, 2011](#)) focus on unethical behavior often stemming from the pressures researchers feel to make their arguments and build their careers. Failure to identify the shortcomings of one's research or to suppress findings of "no difference" may be mildly unethical practices. Taking shortcuts that involve falsifying or distorting data or research methods, or actually hiding negative findings, however, are more egregious violations. Such fraud delays scientific advances and undermines the public's trust in scientific endeavors. Plagiarism, another form of research misconduct, occurs when a researcher claims as his or her own work done or written by others (e.g., colleagues and students) without adequate citation. Although not technically illegal if the "stolen" materials are not copyrighted (e.g., presenting Ibn Khaldun's words as one's own), plagiarism compromises research integrity, which charges scholars to be honest, fair, and respectful of others and to act in ways that do not jeopardize their own or others' professional welfare.

Classification of these behaviors as forms of scientific misconduct derives in part from philosophical principles similar to those underlying the concern for the protection of the welfare of human research participants. [Israel and Hay \(2006\)](#)

analyze philosophical approaches to how people might decide what is morally right—what should be done—in certain circumstances. One approach, focusing on the consequences of a behavior, comes from the writings of utilitarian philosopher John Stuart Mill and invokes a cost–benefit analysis. Essentially, if the benefits that arise from a behavior outweigh the risks or harm associated with that behavior, then it is morally acceptable. This approach, however, begs the question of what constitutes a benefit or harm. In contrast, nonconsequentialist approaches, originating in the works of Immanuel Kant, suggest that what is right is consistent with human dignity and worth. This perspective also emphasizes duties, irrespective of the consequences per se.

Social psychologist Herbert Kelman (1982) emphasizes consistency with human dignity in his evaluation of ethical issues in different social science methods. Kelman notes two components of human dignity: identity and community. The former refers to individuals' capacity to take autonomous actions and to distinguish themselves from others, whereas the latter regards the interconnections among individuals to care for each other and to protect each other's interests. Thus, to promote human dignity requires people to accord respect to others, to foster their autonomy, and to care actively for their well-being. In so conceptualizing human dignity, however, Kelman also draws attention to nonutilitarian consequences: "Respect for others' dignity is important precisely because it has consequences for their capacity and opportunity to fulfill their potentialities" (p. 43). For example, lying to colleagues about scientific results or deceiving subjects about the purpose or procedures of an experiment violates human dignity by creating distrust within a community and/or by depriving individuals of information to meet their needs or to protect their interests. The principle of human dignity, a "master rule" according to Kelman, may be useful in resolving conflicts that arise in the development of a research project by weighing the costs and benefits of taking various courses of action and then choosing the actions that are most consistent with the preservation of human dignity.

Kelman's abstract approach to human dignity substantively undergirds the three more accessible principles promulgated in the Belmont Report (National Commission, 1979), which exists as the cornerstone for the federal requirements for the protection of human research participants. First, respect for persons captures the notion that individuals are autonomous agents and allows for the protection of those with diminished capacity (i.e., members of vulnerable populations with limited autonomy due to legal status, age, health, subordination, etc.). Second, beneficence refers to an obligation to maximize possible benefits and to avoid or minimize potential harms. This principle is consistent with Kelman's emphasis on the means to resolve conflicts between rules by opting for the best means to preserve human dignity. Third, the principle of justice regards who ought to receive the benefits of research and bear its burdens. In this sense, justice pertains to the selection of research participants insofar as those who bear the burden of research should also be the ones to benefit from it.

In addition, the justice principle requires reasonable, nonexploitative procedures. These elements of justice highlight Kelman's emphasis on the community in protecting human dignity.

The principles of the Belmont Report encapsulate the extensive work of bioethicists such as Beauchamp and Childress (2008), who offer “calculability and simplicity in ethical decision making” (Israel & Hay, 2006, p. 18), especially in comparison to the lofty abstraction of other philosophical traditions. The abstract moral principles provide the larger framework for considering what is right and wrong in the pursuit of a scientific understanding of social behavior. In other words, it is important not to lose the fundamental concern with protecting human dignity—both for the individual and for the community—when designing a study, interacting with study participants, and communicating the study’s results.<sup>3</sup> Researchers must consider the ethics of their research and take steps to protect study participants even when they are not strictly required to do so by federal regulations.

### III ETHICAL ISSUES IN LABORATORY EXPERIMENTS

Ethical concerns arise with the use of any social scientific method, both in general terms and in terms of particular characteristics about each method. Laboratory experiments challenge the protection of human dignity in ways like other methods and also in unique ways.

#### A General Ethical Concerns

Methodology chapters on research ethics include lists of concerns (e.g., Babbie, 2013; McBurney & White 2012; Neuman, 2011), varying in length and employing different labels. Four interrelated topics, however, generally emerge: objectification of research participants; potential harms; coercive, exploitative, or intrusive practices; and maintenance of privacy or confidentiality. Each concern implicitly or explicitly raises issues of power and trust between the researcher and study participant. Scholars earn the authority to conduct social research by upholding their ethical responsibility to protect the interests of their participants (Neuman, 2011).

##### 1 *Objectification*

As a consequence of the traditional view of science that separates the observer (the scientist) from the observed, researchers often deem the individuals that

---

3. Although Kelman (1982) draws attention to identity and community as complementary aspects of the principle of human dignity, some (e.g., Hoeyer, Dahlager, & Lynöe, 2005) argue that biomedical research ethics tend to privilege the individual and his or her autonomy over the political implications of the research. The latter concern grows out of anthropological research traditions that share a collective memory of the complicity of some researchers in the execution of power over indigenous peoples and that depend on the development of rapport, reciprocity, and trust with their research informants.

they invite to participate in their studies as their “subjects.” The term “subject” transforms an autonomous actor or moral agent into an object of study. The researcher controls—manipulates in some instances—the context in which the subject is allowed to behave. The extent of control is variable, ranging from very little (e.g., observation studies occurring in public places and mail or online surveys) to a great deal (e.g., laboratory experiments). Use of the term “subject” compared to the more neutral term “study participant” may also raise a graver consequence than the potential loss of autonomy.

The transformation of individuals into subjects carries the danger of treating participants as merely “research material” (Veatch, 1987) and separating the researcher’s humanity from that of his or her study participants (Neuman, 2011). Conceiving of subjects as objects rather than people, researchers distinguish themselves as members of a more powerful group and subjects as members of a subordinate group. In doing so, researchers may grow callous with regard to the potential harms that their studies inflict on their subjects.

An alternative view of the role of research participants, embraced by feminist and humanistic social scientists, is that of collaborator (McBurney & White, 2012). Some scholars take this conception to an extreme by encouraging participants to contribute to the study design. Others employ a more general prospective participant perspective by seeking ethical advice from members of groups who will serve as study participants (Fisher & Fyrberg, 1994). It is critical for scientists to be mindful that experimental participants are human beings deserving esteem and respect in line with Kelman’s (1982) principle of protecting human dignity.

## 2 Potential Harms

A study participant may or may not realize that he or she is being treated as an object. If such a realization occurs, the person may feel distressed or resentful. Whether the level of those negative feelings exceeds similar feelings experienced in daily life, however, may depend on other elements of the situation, such as the researcher’s provision of information and reactions to any display of distress. Generally, in assessing the potential harms of a research study, concerns pertain to: (1) the nature of the harm; (2) the degree or intensity of the harm, often assessed by comparing with harms associated with activities of everyday life; and (3) the actual likelihood or risk that a harm will occur, distinguishing between the possibility and probability of a harm (i.e., many harms may be possible, but the probability of any particular harm may be extremely low).

Experiences that detract from well-being or create an ill effect constitute harms (Kelman, 1982). Sieber (2003) identifies various types and degrees of harms arising in social science research. She also offers means to ameliorate each type of harm.

The most extreme type of harm, physical, is very rare in social and behavioral research (Neuman, 2011) compared to the risk of it in biomedical studies. Sieber (2003) describes physical harm as minimal (transitory or a minor injury)

or major (involving assault or creating a life threatening situation). Researchers must take appropriate safety measures, including anticipating the possibility of risk, screening out vulnerable populations, and monitoring participation.

More common in social science research studies is inconvenience, psychological harm, or social harm. Inconvenience may stem from asking participants to engage in boring, repetitive tasks or asking for the time of individuals with many existing demands. Generally, forewarning of and consent to what a study involves ameliorates such harm.

Psychological harm is, perhaps, the most likely risk in social and behavioral investigations. Researchers sometimes create situations that lead to embarrassment, worry or anxiety, depression, shame or guilt, or even loss of self-confidence or self-esteem. These discomforts may stem from requesting that participants reveal private facts or traumatic experiences or act in ways they did not anticipate doing. In doing so, individuals may come to question aspects of their own identities. The degree to which subjects experience these harms is highly variable, depending on the characteristics of the research study (i.e., the questions asked, the conditions manipulated, the tasks required, and employment of deception) and the participant him/herself. [Flagel, Best, and Hunter \(2007\)](#) provide evidence that minimal-risk psychological studies typically induce little stress, unless participants worry about confidentiality or performance evaluation. Researchers can reduce the possibility of stress by affording study participants respect by giving details about a study (including the possibility of these psychological harms), ensuring confidentiality, and capitalizing on any psychological benefits a study might provide (see [Decker, Naugle, Carter-Visscher, Bell, & Seifert \(2011\)](#) on how study participation was helpful to victims of childhood abuse). It is incumbent upon the researcher to sense distress in participants and to reassure them by giving them the opportunity to skip questions or discontinue participation.

Social harms generally refer to threats to an individual's reputation or his or her relationships with others. Revealing private facts about a study participant (e.g., HIV/AIDs status or drug use) may compromise his or her standing in a workplace or community. Subjecting a person to an embarrassing situation may decrease the level of esteem accorded to him or her by others—friends, family members, co-workers. Careful consideration of the likelihood that study procedures may cause these harms and how they may be circumvented are tasks of the researcher. Also, as discussed later, to avoid compromising a study participant's standing, the researcher must be prepared to ensure the confidentiality of the data collected.

[Sieber \(2003\)](#) also describes economic and legal harms. Economic harms range from the loss of a few dollars to the loss of financial opportunities, credit, jobs, and so forth. Again, it is important for the researcher to ensure confidentiality and potentially to compensate the participant should loss occur. The potential for legal harm is most likely in studies of illegal behavior and includes involvement with lawyers or law enforcement and the potential for convictions.

A key strategy to protect participants from such harms is the assurance of anonymity and the *a priori* recognition of the risks both for the participant and for the researcher (should he or she witness illegal activities and/or be called upon to testify).

As [Babbie \(2013\)](#) notes, just about any study might potentially harm someone in some way. The strongest scientific grounds must exist for pursuing any study that causes participants to suffer ill effects. The researcher is responsible for recognizing and minimizing risks of harm.

### 3 Coercive, Exploitative, and Intrusive Practices

Potential harms are typically consequences of the researcher's actions. Equally harmful, however, are coercive, exploitative, or intrusive practices that investigators may knowingly or unknowingly employ in their quest for knowledge. The ethicality of this sort of practice highlights the power relationship between the investigator and potential participants.

To the extent that a researcher has something (e.g., a treatment, money, a grade, or other special rewards) that a potential participant might desire, he or she is in a power-advantaged position vis-à-vis those who might enroll in the researcher's study. Using the desired resource to entice participation may constitute a form of coercion. To avoid the appearance of or actual coercion, emphasis on such enticements must be minimal—that is, researchers should limit the amount of money to be paid to the subject and must never make course grades dependent on study participation. [Singer and Couper \(2008\)](#) examined the issue of incentives on survey participation. They found that “larger incentives induce greater participation than smaller ones, for larger as well as smaller risks, and larger risks induce less participation than smaller ones do” (p. 53). Importantly, that larger survey incentives alone do not entice people to take on larger risks might quell some concerns about coercion. [Ripley \(2006\)](#) offers some guidelines on paying research participants. Generally, researchers must understand that participants are volunteers who are—and who must know that they are—free to decline participation.

Even once individuals have agreed to participate, researchers cannot exploit this agreement by asking them to answer questions or perform behaviors that had not been previously described or by prohibiting their withdrawal from the study. Although people may be apprehensive about participating in studies, once committed, they may feel obligated as “research subjects” to stick with it, despite experiencing discomfort ([McBurney & White, 2012](#)). To take advantage of adherence to such role demands is a form of power use, of exploitation. To minimize the potential for exploitation, researchers must be mindful of the voluntary nature of subjects’ participation as well as recognize their obligation to respect their participants.

Such respect may also temper how intrusive the requirements of a study may seem to a participant. [Babbie \(2013\)](#) notes that any study takes time and energy, and some ask participants to reveal information about which even friends or

co-workers know nothing, even in the absence of direct benefits to the participants. Researchers, however, may recognize the need to obtain certain types of information to advance knowledge or to develop policies to address social problems. Even though the potential to generate social benefits may justify such intrusive practices, the researcher who wields expert power must fully inform participants about why particular types of information are being solicited or types of behaviors observed. With such explanations, the individual may then decide whether or not to participate. The voluntariness of participation is one means to equalize the power relationship.

#### *4 Maintenance of Privacy and Confidentiality*

Insofar as researchers implicitly or explicitly ask their study participants to disclose about themselves or their behavior, they are obtaining information that people might want to keep private. Ethical concerns arise regarding how those revelations will be held. Promises of anonymity or confidentiality are a basis for protecting individuals' privacy (Babbie, 2013).

Normative rights to privacy assure individuals that they need not reveal information about themselves. In other words, people control who has access to such information, which in turn affects their willingness to participate in research (Sieber, 1992). Moreover, privacy allows individuals to establish personal boundaries. When individuals agree to participate in a study, they provide the researcher with access to certain types of information. In some instances, it is possible to assure research participants that their revelations will be anonymous—no names or other unique identifiers are associated with the information given.

Typically, however, researchers know their subjects' names and may be able to link those names to their revelations. Such links allow researchers to track study participants for follow-ups or triangulation of data sources (e.g., official records with interview or survey responses). To protect the privacy of study volunteers, researchers may promise that the data will be held confidentially. Specifically, confidentiality refers to how records about the study participant will be handled (Sieber, 1992), including identification of who may access the information, how the information will be securely stored, and how it will be disseminated. Potential for breaches of confidentiality is a major source of possible harms in social and behavioral research studies (Citro et al., 2003). Technological developments in data collection, processing, dissemination, and analysis may compound the likelihood of social, legal, and economic harms. For example, the responses of participants in online experiments or surveys may be traceable to individuals, if only via IP addresses, and stored data may be accessible to unauthorized personnel.

To prevent breaches of confidentiality and thus protect private information, researchers may enact several interrelated strategies (Babbie, 2013), including the ethical training of researchers and their assistants, de-identification of study documents by removing names, separate storage of and limited access to master files linking identities with study data, password-protected computerized data

files, and presentation of data only in aggregate form or only with pseudonyms. Although researchers usually stress that they will keep data confidential and do all that they can to avoid public disclosure of identified information, they can only do so to the extent allowed by law. Law enforcement officials may, under certain conditions, subpoena research documents and failure of researchers to provide them may result in contempt charges. A certificate of confidentiality issued by the U.S. Department of Health and Human Services (via the National Institutes of Health or the National Institute for Justice) protects researcher-participant privilege and may be granted for funded or unfunded studies for which protection of confidentiality is necessary to achieve study purposes (Sieber, 1992). Researchers may obtain such certificates for studies on sensitive topics (e.g., mental health, use of drugs or alcohol, sexual practices, and illegal behavior) that might result in social, economic, or legal harm to the study participants should identifiable information be released.<sup>4</sup>

Most experimental investigations are likely to fall outside of the range of studies that qualify for certificates of confidentiality. Thus, it is imperative that experimenters protect study participants' identities in other ways. By sincerely informing research participants that the information that they provide will be held confidentially, the researcher indicates respect, potentially guards against various types of harms, and minimizes perceptions of objectification and power-based practices. Researchers who are cognizant of these ethical issues and act on them in all stages of research, from recruitment of participants to data collection and presentation of results, are likely to reinforce the public's trust in the research endeavor in general. Experimentalists must also attend to several specific issues.

## B Specific Ethical Concerns

Experimental research is unique from other methodological approaches because investigators' emphasis on the isolation of key factors (and thereby the control of extraneous factors) leads them to use deception to enhance experimental control. Even when deception is not used, there is a tendency to not fully inform study participants about the study to develop more control over experimental treatments. In such cases, the completeness of informed consent becomes an issue. Also, because laboratory studies may involve special equipment located in dedicated spaces, experimenters have come to rely on participants who are readily available, such as those who are enrolled in courses within departments.<sup>5</sup> In some instances, students become part of highly organized "subject pools," the members of which take part in research studies as part of the curriculum.

---

4. For more information, consult <http://grants2.nih.gov/grants/policy/coc>.

5. For experimental research, the use of nonrandom samples, such as a pool of college students, is unproblematic as long as a study is theoretically based and the college students meet the theory's scope conditions (see Webster and Sell, Chapter 1, this volume)

Concerns about deception and subject pools embody issues of objectification, potential harms, coercion, and confidentiality in various ways.

## 1 Deception

Research involving deception has a long history, especially in psychology (Korn, 1997). Deception refers to acts of providing false information or withholding information intended deliberately to mislead others into believing something that is untrue (Bordens & Abbott, 2013; Sieber, 1992; Sieber, Iannuzzo, & Rodriguez, 1995). Such acts are distinct from the typical practice in experimental research of not fully acquainting subjects in advance with all aspects of the research being conducted, such as the hypotheses to be tested and the nature of all of the experimental conditions (Baumrind, 1985; Ortmann & Hertwig, 2002). Use of deception research is hotly debated largely due to its ethical implications (e.g., see Baumrind (1985), who opposes deliberate deception, and Bonetti (1998), who argues its benefits). Yet insofar as the nature of deception and the nature of the research in which it is embedded vary, the practice continues, albeit with precautions to prevent injury to human dignity.

A number of scholars have noted the potential harms of deception research at all levels: study participants, researchers, researchers' profession, and society (Elms, 1982). At the individual level (Baumrind, 1985; Korn, 1997; Sieber, 1992), deception undermines research participants' autonomy and right to self-determination because the lack of full disclosure of experimental purpose and procedures impairs participants' decision-making ability. Individuals may also feel embarrassed or even suffer a loss of self-esteem upon realization that they have been duped. Such consequences deprive research participants of the respect that ensures human dignity in interpersonal relationships, which may enhance objectification of the subject (Kelman, 1982). In addition, the failure of researchers to demonstrate respect and veracity in their dealings with study participants may undermine their own reputations in the long term.

The cumulated effects of the use of deception may negatively affect a profession as well. Researchers within a discipline who rely heavily on deception methodologies may be seen to abuse the power associated with their expertise (Korn, 1997). In addition, deception may stimulate suspicion and negative attitudes about research among potential subjects that undercuts the work of their colleagues using nondeceptive methods (Baumrind, 1985; Bordens & Abbott, 2013; Sieber, 1992). Economists have a related practical concern: if study participants are aware that deception may be used, their behavior may be shaped more by psychological reactions to suspected manipulations rather than actual situational circumstances (e.g., induced monetary rewards) (Davis & Holt, 1993). Economists (Davis & Holt, 1993; Hertwig & Ortmann, 2001, 2008) also emphasize the importance of maintaining a reputation for honesty and claim that the use of deception may damage that reputation, even if practiced by only a few. More generally, the concern is that deception in research may erode the public's trust in social scientific endeavors (Baumrind, 1985; Kelman, 1982;

Korn, 1997) and lead to cynicism (Neuman, 2011). This erosion in trust may play out in a university by making students and others unwilling to volunteer for all types of research. At the societal level, experience with deception in research may create a willingness in others to use deception and contribute to a lack of trust in general (Elms, 1982; Sieber, 1992), thereby eroding a fundamental social value that binds relationships (Kelman, 1982).

Ortmann and Hertwig (2002) empirically assess these potential harms. Drawing largely from psychological studies, they show that study participants' direct experience with deception generates suspicion, which in turn may alter their judgment and behavior in subsequent studies. Thus, use of deception may invalidate the results of future investigations potentially through the operation of demand characteristics that lead participants to guess at study objectives and to alter their behavior accordingly. Little empirical evidence exists for the spill-over effect of indirect experience—that is, knowledge that deception is used in laboratory studies—on emotional or behavioral responses to research.

Whether deception causes direct harm to subjects is ambiguous, given the conflicting evidence (Ortmann & Hertwig, 2002). For example, Christensen's (1988) review of the literature demonstrates that participants in deception studies do not perceive that they have been harmed. Likewise, although Epley and Huff (1998) detect the direct effect of deception on suspicion, few negative reactions to being in a deception experiment emerged. Similarly Boynton, Portnoy, and Blair (2012) confirm that prior experience with direct deception makes study participants suspicious of experimental cover stories, but they are not particularly bothered by their experience with deception. The lack of experienced negative reactions contrasts with findings from Fisher and Fyrberg (1994) showing that when directly asked how they are likely to feel upon learning that they participated in a study involving deception, college students indicate anticipating negative feelings such as discomfort, sadness, or embarrassment.

Little evidence exists for the potential threats to the credibility of a profession or threats to human dignity in society (Ortmann & Hertwig, 2002). Student evaluators of deceptive research (Fisher & Fyrberg, 1994) largely believe that studies are scientifically valuable and valid and that a study's social benefits outweigh the costs. The potential harms signal that deception should be used cautiously and only with substantial scientific and ethical justification.

Typically, the use of deception in research is defended in terms of its benefit to scientific ends, which in turn benefit society (Cook & Yamagishi, 2008; Elms, 1982; Korn, 1997). Also, some types of studies (e.g., studies of false expectations) are difficult, if not impossible, to perform without deception (Sell, 2008). These benefits are weighed in view of the costs or potential harms identified previously. Sieber (1992; Sieber et al., 1995) offers four defensible justifications for deception research. First, deception may be employed to achieve stimulus control when no other alternative procedures are feasible to ensure the execution of valid research with scientific, education, or applied value and not simple trickery. Second, deceptive conditions may allow the study of responses

to low-frequency events. Third, designs involving deception must not present any serious risk of harm to study participants. Fourth, deception may be an appropriate means to obtain information that would otherwise be unobtainable because of subjects' anxiety, fears, embarrassment, defensiveness, and so forth. These considerations diminish a researcher's ability to coerce subjects into a study involving unjustified deception. Decisions about the ethics of deceptive research extend beyond these justifications. As [Ortmann and Hertwig \(2002\)](#) suggest, "whether deception ... is considered acceptable by a participant, is a function of ... the nature and severity of deception, the methods of debriefing, and the recruitment mode" (p. 117).

Some deceptions are relatively innocuous, involving false expectations about experimental procedures, whereas others are more serious, such as providing subjects with false feedback about their performances on tasks that bear on their self-evaluations outside of the laboratory ([McBurney & White, 2012](#)). Deception may be active (e.g., misrepresenting the purpose of the research, misidentifying the researcher, providing misleading information on equipment or procedures, and using confederates of the experimenter) or passive (e.g., concealed observation and unrecognized conditioning) (see [Bordens & Abbott, 2013](#)).

[Sieber et al. \(1995\)](#) codify the harmfulness of deception by: (1) the kinds of failures to inform (e.g., false informing about a study or its procedures, no informing, consent to deception, and waiver of the right to be informed); (2) the nature of the research in terms of the harmfulness or privacy of the behavior, the confidentiality of the data collected, and the power of the means of inducing the behavior; (3) the topic of the deception (e.g., the study's purpose, stimulus materials, feedback about self or others, and actual participation in a study); and (4) the extent of "debriefing" or explanations given after the study. Deception studies focusing on harmful behaviors that people would most likely prefer to keep private, induced by powerful means, involving false information about oneself without suitable debriefing are most likely to (rightfully) prompt ethical objections. Sieber et al. conclude that "to be acceptable, deception research should not involve people in ways that members of the subject population would find unacceptable" (p. 83) or cause discomfort at a level that would have prevented people from participating had they fully known in advance what the study involved ([Fisher & Fyrberg, 1994](#)).

[Pascual-Leone, Singh, and Scoboria \(2010\)](#) offer a list of items to consider when potentially planning a study using deception. Use of the checklist ensures that researchers have thoroughly thought through potential other means to test their ideas, have recognized necessary precautions if deception is used, and have carefully crafted the nearly always necessary debriefing.

One way researchers can assess whether study volunteers might find deception objectionable or harmful is to debrief pretest participants and then modify procedures to reduce or ameliorate potential harms and clarify important debriefing information. For both pretest and study participants, debriefing involves dehoaxing and desensitizing ([Holmes 1976a, 1976b](#)). Dehoaxing is

the process of informing subjects after the session of the experiment's true purpose and revealing the nature of the deception. Indicating how equipment actually worked or providing actual performance scores on tests is part of this process. The researcher needs to dehoax without increasing subjects' embarrassment over being duped and with the intent of eliminating any mistrust engendered by the study with regard to the scientific endeavor. [Fisher and Fyrberg \(1994\)](#) find that if people feel embarrassed after learning of the deception, they are unlikely to reveal that emotion to the experimenter but might rather reveal annoyance or anger.

Desensitizing focuses on attempting to remove any emotional harm (e.g., discomfort, anxiety, or distress) that the study and specifically the deception may have caused. The intent is to restore a sense of positive well-being, presuming that any emotional distress is temporary; studies that damage self-esteem or involve lasting harms should be avoided in the first place ([Elms, 1982](#); [Sieber, 1992](#)). If necessary, the researcher should draw upon additional resources to handle the stress created by the study as well as, potentially, by the debriefing. By providing the participants with the opportunity to ask questions, receive adequate responses, and withdraw from the study, the researcher may further minimize negative consequences of the study. [Oczak and Niedźwieńska \(2007\)](#) incorporate an educational procedure into their debriefing that provides participants with "insight into relevant deceptive practices and how to recognize and deal effectively with them" (p. 49). As a result, their study participants expressed a more positive mood and attitudes toward research. These actions demonstrate respect for study participants and may engender continued trust in scientific research.

Deception research continues in some disciplines (e.g., psychology, sociology, and political science) and is de facto prohibited in others (e.g., economics). [Geller \(1982\)](#) suggests that forewarning of the possibility of deception, but not giving the actual description of it, may be an alternative, more ethical approach. However, associated methodological issues (e.g., demand characteristics and threats to random sampling) may compromise the validity of the data. [Fisher and Fyrberg \(1994\)](#) find that potential study participants think that forewarning would discourage participation due to the discomfort people may feel in being "controlled."

Social scientists continue to debate the use of deception, often invoking different philosophical approaches to considering what is ethical. At a minimum, a focus on protecting human dignity may circumvent the emergence of potential harms to individuals, to professions, and to society. Moreover, careful pretesting may alert researchers to unexpected harms and lead them to devise more ethical procedures. Ironically, to the extent that experimental situations avoid mundane realism yet appear realistic enough to be taken seriously, the potential harms may be less likely to arise because subjects may attribute deception to the artificiality of the laboratory rather than to the specter of having been treated without dignity.

## 2 Subject Pools

To provide a ready stream of research subjects, many (psychology) departments require students in introductory or other lower division courses to participate in studies conducted by faculty or other (graduate and undergraduate) students (see Sieber & Saks, 1989). The number of hours of participation or number of studies in which a student must be a part varies, but the participation counts toward the course grade. In other words, students get credit for becoming research participants or they are offered alternative activities (e.g., writing a paper, doing extra homework, taking a quiz, and viewing a movie or demonstration) to earn the same credit.

The creation and use of such subject pools produces conflict between the development of knowledge and the ethical treatment of research participants (Sieber & Saks, 1989). To the extent that students feel pressured into participation, they suffer from both objectification (i.e., existing as fodder or “raw material” for someone’s study) and coercion. In effect, researchers violate subjects’ autonomy and demonstrate disrespect, thereby compromising human dignity.

As previously noted, voluntariness in the decision to participate in a research study is a key aspect of ethical treatment. In psychology, investigators support the long-standing and pervasive practice (Landrum & Chastain, 1999; Sieber & Saks, 1989) by emphasizing the benefit of subject pools: the development of valid scientific knowledge while providing students with an important hands-on educational experience. The educational value of participation in research via the subject pool is students’ direct exposure to the research domain and the process of scientific inquiry. In addition, some argue that students may learn how to conduct research ethically by participating in presumably ethical experiments.

Some studies show that students who participate in research hold more favorable attitudes toward it than those who do not participate (Waite & Bowman, 1999) and reflect positively on the significance of the research or what they learned (Moyer & Franklin, 2011). In contrast, in largely undergraduate departments, students do not typically assess the educational value of their research participation (Landrum & Chastain, 1999). Also, it is difficult to assess the educational value of participation in studies in departments hosting graduate programs because of the wide variation in the mechanisms to ensure a valuable educational experience (e.g., from providing an abstract of results months later to lengthy discussions of the study coupled with the subjects’ own evaluations of the research experience) (Sieber & Saks, 1989). To contend that required participation in research has educational value, Sieber (1999) urges researchers to debrief participants verbally and in writing, without jargon, on the background of the study and the current purposes immediately following the research session.

In the absence of a certain educational benefit, the potential to violate the principle of voluntary participation jeopardizes the ethicality of subject pool practices. Although provision of alternative activities for the same credit is one means to ameliorate the appearance of coercion, the typical alternative of

writing a paper may be more onerous ([Sieber & Saks, 1989](#)). In addition, coercion may arise in the debriefing process. [Fisher and Fyrberg \(1994\)](#) note that students may feel restrained in the questions they ask or what they say once they learn about the study for fear that saying the wrong thing might jeopardize their credit. In effect, the students may not feel totally free to withdraw from the study because they may fear—rationally or irrationally—losing the required credit (even though the consent document outlines voluntariness and conditions for receiving credit). Thus, it is important for researchers to reassure subjects that they will receive credit, regardless of their responses during the debriefing or even if they withdraw partway through the study.

[Sieber \(1999\)](#) and [Landrum and Chastain \(1999\)](#) suggest procedures to demonstrate respect toward members of the subject pool, thereby ensuring ethical treatment. For example, information on subject pool requirements should be widely available to students (via course Web sites and descriptions, specific handouts phrased clearly and respectfully, etc.). Students should also be made aware of the process by which they may file complaints about their experience in research studies. Also, alternatives to research participation should be equivalent in terms of educational experience, time, effort, and so forth. Departments may also devise means to monitor the behavior of researchers with regard to their awareness of subject pool protocols and the proper treatment of participants in their studies. Despite the long existence of subject pools, actual practices must be in keeping with professional codes and the federal regulations.

## IV MEETING REGULATORY REQUIREMENTS

In the United States, the federal regulations for the protection of human research participants (45 CFR § 46) derive from the ethical principles developed in the Belmont Report (1979): justice, beneficence, and respect. Adopted by 15 federal departments and agencies, these regulations constitute the “Common Rule” governing human subject research sponsored by the federal government. The regulations specify mechanisms to ensure IRB review of research, informed consent of study participants, and institutional assurances of compliance ([Dunn & Chadwick, 2004](#)). Most universities extend the review to all funded and unfunded research conducted by faculty, staff, and students and instigate additional criteria to be met prior to the approval of a project. Debate centers on whether the federal regulations fulfill the promise of protecting human dignity, especially in reference to social and behavioral research. Given the moral mandate for ethical research, researchers must learn to navigate the regulatory system.

### A IRB Review of Social Science Research

Presumably, IRBs protect the human dignity of individuals who volunteer to participate in research and ensure compliance with federal regulations. Some

commentary, however, suggests that IRBs may lose sight of the ethical foundations of human research protections. DuBois (2004) notes that many education programs focusing on human protections offer inadequate reasons for compliance with regulatory demands. Ironically, promoting compliance for its own sake or to avoid potential penalties is contrary to moral notions of promoting human dignity. He also suggests that when IRBs unnecessarily impede research, they are acting unethically. Indeed, DeVries et al. (2004) show that researchers frequently view IRB review as an unnecessary bureaucratic hurdle circumventing their presumed rights to conduct research with human subjects. Such a view fails to recognize the IRB's ethical mission and the tenuousness of their right to pursue research with human subjects (Oakes, 2002). Like access to grants and publishing outlets, researchers face different types of reviews. In addition to their responsibility to ensure human well-being, IRBs also have a responsibility to foster research (DuBois, 2004) and not simply to mindlessly enforce regulations without attention to the diversity of research questions, subject populations, and methods (Eckenwiler, 2001).

Social scientists have long been among the most vocal critics of IRBs (see Citro et al., 2003). Designed largely for biomedical research, the regulations give short shrift to social and behavioral research (Oakes, 2002; Singer & Levine, 2003). Social scientists fault IRBs for being “overprotective, unreasonable in their demands for consent, impractical in their directives for the protection of confidentiality, and excessive in the time required for review” (DeVries et al., 2004, p. 352). These failings, they contend, stem from three interrelated sources.

First, although some institutions do have boards consisting wholly of social scientists, too often too few social scientists serve on IRBs. In the absence of social scientific knowledge on the boards, the second source of problems is the application of a biomedical model of research to social and behavioral investigations, which results in unreasonable requirements or in overestimating risks of nonphysical harm and underestimating the benefits of a project because the study is unlikely to produce direct benefits (e.g., improved health). For example, it is certainly unreasonable to obtain written documentation of informed consent from illiterate study participants and may be incongruous to demand parental consent for adolescents' participation in a minimal risk study that concerns behavior that parents often forbid or about which they are unaware (e.g., smoking) (see Diviak, Curry, Emery, & Mermelstein, 2004). A third problem, documented by the direct observation by a qualitatively trained sociologist (see DeVries et al., 2004), compounds the unintended consequences stemming from the inappropriate application of the biomedical model: IRBs are likely to scrutinize more intensely social and behavioral research protocols than biomedical ones despite the far lower risks of harm.

Failures of IRBs to understand the nature of social and behavioral research and to use the flexibility inherent in the federal regulations gave rise to emphasis on regulatory compliance rather than the execution of a reasonable approach

to protecting the human dignity of individuals involved in social scientific research. Scholars offered a number of means to rectify these problems, including having social scientists grow familiar with federal regulations, serve on IRBs, and secure ethics training; and encouraging IRBs to learn more about issues associated with social and behavioral research (Citro et al., 2003; DeVries et al., 2004; DuBois, 2004; Oakes, 2002). As a result, in recent years, many IRBs have grown more sensitive to the nature of different types of studies and the best ways to use regulatory flexibility to ensure appropriately the well-being of social research participants. In effect, healthy debate and communications between researchers and IRBs promulgated changes to address some criticisms.<sup>6</sup> Nonetheless, scholars must learn how to navigate the IRB review process at their specific institutions.

## B Navigating the IRB

Although institutions have different procedures for obtaining IRB review, all are guided by the federal regulations and advisory postings of the OHRP (see footnote 2). Researchers and IRB members and staff have distinct sets of responsibilities (see Dunn & Chadwick, 2004).

Researchers bear the ultimate responsibility for the welfare of their study participants (Table 2.1). Consent procedures and IRB review help to ensure human subject protections, but it is incumbent upon the researcher—the expert and the contact for study participants—to anticipate threats to well-being, monitor the study for the emergence of harms, cease data collection until changes can be made to minimize the risk of harm, and notify the IRB of any adverse events. Researchers must use their professional judgment to balance the benefits

**TABLE 2.1 Researchers' Responsibilities in Protecting Research Participants**

Ensure the safety and welfare of research participants

Use good professional judgment to determine research benefits in view of potential harms of varying types

Demonstrate knowledge about federal regulations and institutional policies

Comply with federal regulations and institutional policies

Convey information to research participants in a way that they can understand

6. Based on new guidance from the OHRP and workshop offerings at professional meetings for social and behavioral research, the IRB Director at Emory University communicated to me her observation that many IRBs have attempted to address the (not unfounded) criticisms voiced in earlier works cited previously.

of their research in view of risks of harm or violation of rights and familiarize themselves with federal regulations and institutional policies. They may demonstrate compliance with the requirements by clearly communicating to the IRB the substantive questions of their work and the methodology to be used; outlining the benefits of the research in view of the likelihood of various potential harms; and indicating how they will obtain the consent of study volunteers and maintain the confidentiality of their responses. Moreover, researchers must be able to convey similar information to study participants in a readily understandable way. In effect, researchers mesh the ethical concerns in their work with the demands of regulations and institutional policies.

The IRB has the responsibility of guiding researchers through the review process. IRB members and staff must clearly indicate what information must be submitted and readily answer investigators' questions. Those answers, moreover, should take into account the nature of the research and the subject population to avoid a cookie-cutter approach, which may create unintended negative consequences. IRB reviewers must have sufficient expertise to address the wide range of substantive issues addressed in the studies that they review. The boards include community representatives who catch harms to which professional researchers may be blind. IRB reviews should unfold in a reasonable amount of time and provide explanations for requested changes. By doing so, IRBs further educate investigators.

The federal regulations identify three types of IRB review (see [Citro et al., 2003](#); [Oakes, 2002](#); and the Web site in footnote 2). Regulatory definitions of research and of human subjects determine which studies are excluded from IRB review. Exclusions include studies not involving direct (e.g., interviews) or indirect (e.g., surveys) interaction. Most social scientific research falls into the “minimal risk” category, in which the probability and magnitude of harms are no greater than those ordinarily encountered in daily life and may be ameliorated by researchers’ actions (e.g., stopping a distress-inducing procedure and ensuring data confidentiality). Institutions, not researchers, judge the level of risk and designate the review type: exempt, expedited, or full board.

When studies entail public observation of behavior, anonymous surveys, or interviews with adults who are not vulnerable in any way (e.g., due to pregnancy, diminished capacity, or incarceration), they usually involve only minimal risk and fall into the “exempt” review category. Although, technically, such studies fall outside of the purview of the federal regulations, most institutions require the IRB to designate the study as “exempt.” Minimal risk studies with written documentation of consent, possible risk of loss of confidentiality, collection of voice, video, and/or image recording for research purposes, or collection of small biological specimens through noninvasive clinical procedures are likely to be subject to “expedited” review. Institutions often similarly execute exempt and expedited reviews, employing rolling reviews by a committee or staff member. IRBs reserve full board review for studies involving greater than minimal risk, vulnerable populations (e.g., children, prisoners, pregnant women when

there is a threat to the fetus, and individuals with diminished capacity), and/or deception. At some institutions, changes in or waivers of informed consent elements may also attract full board review. Although researchers sometimes request a particular review type, ultimately the IRB determines the review level. Attempts to thwart higher levels of review are typically unsuccessful. Thus, it is in the investigator's best interest to prepare a thorough, explicit, and honest IRB proposal.<sup>7</sup>

### *1 Preparing an IRB Proposal*

Although most researchers deem the IRB proposal a bureaucratic hurdle to be overcome to secure grant monies and/or collect data, the moral mandate for ethical research should implicitly facilitate the preparation of materials. When investigators cooperate with their IRBs, mutually shared goals of protecting human research participants are likely to be achieved. To ensure success in the IRB review process, investigators must prepare clear, consistent, and complete documents that convey their concern with designing procedures to protect the well-being of study participants (see [Gillespie, 1999](#); [Oakes, 2002](#); [Sieber, 1992](#)).

[Table 2.2](#) describes the documents to include in any IRB application, as well as those especially relevant to experimental studies. Accompanying an institutional application form, the IRB proposal typically includes the research protocol, consent documents, recruitment materials, and instruments. For experimental research, additional items may include the debriefing statement, details of experimental procedures, and specific rationale for any deceptive practices. Institutions may also require signatures from department chairs and/or faculty mentors, statements of researchers' qualifications, and/or documents supporting the research from those with authority over the research context (e.g., clinic administrators and school superintendents or principals). In the case of international studies, researchers may also have to provide information about a local contact or in-country ethics review procedures.

In language appropriate for an educated lay audience, the research protocol describes the substantive question to be studied, situates the project in the larger literature, and demonstrates how the study fills a void and will provide social benefits through augmenting knowledge. The latter ensures that the researcher conveys concern with the ethical principle of beneficence—maximizing good outcomes for science, humanity, and study participants ([Sieber, 1992](#)).

Within the protocol, the researcher should provide detailed information about who will be involved in the study. Characteristics of the participants, their selection and recruitment, and ultimately how they are treated in the course of the study provide the means to assess the principle of justice. Materials indicating how subjects are to be recruited (e.g., flyers, advertisements, Web sites, and

---

7. To intentionally omit relevant information in order to secure, for example, an exempt review would constitute unethical behavior, which would reinforce the views of some IRB staff that investigators are not to be trusted.

**TABLE 2.2 IRB Application Materials**

IRB application (electronic or paper)
Research protocol: describe or identify
Theoretical background of substantive issue to be addressed
Benefits of the research
Research design
Characteristics of study participants
Recruitment procedures, including incentives
Potential types of harms and their likelihood of occurring
Deceptive practices, if any, and how the study meets the requirements for waiver of fully informed consent
General study procedures
Data handling procedures, especially with regard to confidentiality
Data analysis plan
Consent documents
Required elements
Optional elements
Institutional requirements
Recruitment materials (e.g., flyers, ads, verbal pitches)
Data collection instruments (e.g., surveys, questionnaires, measures)
Experimental procedures (e.g., verbal or written instructions to participants, computer screens)
Debriefing materials

scripts) must provide a brief description of what is involved in the study and who the investigators are without overstating the benefits of the study or suggesting excessive inducements, which raise concerns with unjust exploitation or coercion. The application of the principle of justice in experimental work also requires that researchers explain why they plan to use particular sets of participants and how they will make sure that the conditions to which subjects are randomized carry similar levels of risks of harm and of benefits.

The researcher's analysis of the risks inherent in the research should accompany the description of the actual study procedures (e.g., experimental manipulations and explanations of the means to measure dependent variables). Although physical harms are generally unlikely, researchers should assess the

extent to which psychological, social, or other harms may occur in the course of a study. An outline of the means by which to maintain the confidentiality of the records should convince the IRB that the researcher has minimized the potential for a breach in confidentiality, the major risk in social and behavioral research. More generally, discussion of strategies to ameliorate potential harms is part of the risk–benefit analysis and ensures beneficence. Such analysis is particularly important in research involving deception. The process of debriefing and the concerns that may arise in the course of debriefing must be detailed. The researcher should anticipate any negative effects from the debriefing, such as inducing a feeling that the participant is gullible or easily swayed by others’ opinions. In all cases, researchers should assure the IRB that study participants, once fully informed of the nature of the research, have the opportunity to drop out with no loss of benefits (i.e., credit or compensation).

A key element of any IRB proposal is the description of the informed consent process, which is the researchers’ basic means to convey respect to potential participants. Respect signals investigators’ concerns with the autonomy of the participants—their ability to make their own decisions.<sup>8</sup> In this process, the researcher communicates sufficient information to and answers questions from potential study participants so that they can choose whether or not to volunteer for the study (Dunn & Chadwick, 2004; Sieber, 1992). This information exchange occurs before, during, and sometimes after the study. The researcher must provide information at a level that the potential participant may comprehend. The consent process may be fully oral or involve written documents that may or may not require a signature. The default method of informed consent involves a written document that participants review and discuss with the researcher and then sign to signal agreement to participate (“written consent with written documentation”).

What form the consent process takes depends on the nature of the research question, the subject population, and the data collection procedures.<sup>9</sup> Flexibility inherent in the federal regulations allows IRBs to grant waivers of written documentation of consent if the consent form is the only document linking the participant to the study and the principle risk from the study is breach of confidentiality. Under certain conditions, investigators may request waivers of

---

8. The federal regulations recognize that some individuals have limited autonomy and may constitute traditionally vulnerable populations (e.g., prisoners, children, and the mentally disabled). The additional federal safeguards and consent procedures for these populations may also apply to equally vulnerable populations of individuals in subordinate positions or resource-deprived environments (DeVries et al., 2004; Dunn & Chadwick, 2004).

9. Variation in the means of obtaining consent is more common in qualitative or ethnographic research than in survey or experimental research. Citro et al. (2003) note that alternative methods may be more ethical in certain circumstances (e.g., research involving populations whose language skills or cultural attitudes make the default method inappropriate or when research is on illegal behavior or highly sensitive topics and privacy and confidentiality trump the need for written documentation of consent).

other elements of consent. In the case of deception research, an investigator is essentially requesting a waiver of informed consent because the deception makes it impossible to convey all of the details of the study without jeopardizing its scientific validity. To justify the waiver, researchers must explain in their IRB proposal that: (1) the research could not be otherwise carried out without the deception; (2) the study involves no more than minimal risk of harm; (3) omissions in the consent process do not adversely affect participants' rights or welfare; and (4) subjects will be provided with pertinent information whenever appropriate and be given the opportunity to withdraw. In other words, investigators must justify the use of deception in terms set forth by the federal regulations and with recognition of the ethical issues that deception engenders.

The federal regulations also stipulate required elements of consent. The core elements relevant for social and behavioral research include:

- introduction to the study that indicates it is a research project, by whom it is conducted, and the source of funding (if any);
- a description of the study purpose, procedures, and duration;
- disclosure of risks/benefits;
- statements regarding confidentiality of records;
- specification of the voluntariness of participation (which, for example, includes freedom to withdraw from the study or simply to skip questions); and
- the means by which to contact the investigator, IRB officials, and, if relevant, a local contact in international contexts.

Other elements of information should be included if they are relevant to the study and to what the participants need to know to make an informed decision. For example, the number of study participants may affect an individual's decision making if he or she worries about being identified. That situation may be very likely to arise in studies involving subject pools or participation of class members in their instructor's research. Also, if compensated, participants should know in advance what to expect (e.g., a specific hourly amount; a range of possible outcomes, depending on bargaining or exchange behavior; and a specific amount of credit). Compensation may be for participants' time, inconvenience, motives for behaviors under investigation, or the like, but it is separate from the risk–benefit assessment of the study. Institutions also may specify reading level (e.g., eighth grade), structure of written consent document (e.g., with or without headings for subsections, font size, and version dates), phrasing of particular elements, or additional elements of information. Researchers should consult with their local IRBs to learn of additional requirements.

To assess the consistency between what is specified in the IRB proposal and consent documents as well as to gauge the reliability of the researcher's own descriptions of potential harms, investigators must supply copies of their data collection instruments to the IRB. These instruments may be the actual questionnaires, interview guides, or standard measures to which participants will respond. What may seem like harmless questions to an investigator may

raise an IRB reviewer's concerns for the well-being of study participants. The IRB may also consider the explicit and detailed experimental instructions (e.g., what researchers say to subjects to manipulate variables and what information appears on computer screens) to constitute data collection instruments. Such information allows reviewers to imagine what a participant would experience, which may aid them in assessing whether the investigator has anticipated potential risks of harms and/or created procedures to minimize such risks. The detailed description of the study significance and methods permits reviewers to assess the study's scientific merit in view of its risks and benefits.

The IRB proposal is, at one level, much like writing the methods section to a journal article: investigators describe what they intend to do and how they intend to do it. In contrast to simply documenting what was done for others to replicate a study, the IRB proposal demands that researchers consider the implications of their procedures in terms of the rights and well-being of the people they invite to participate in their study. By ensuring justice, minimizing risks, and demonstrating respect in their interactions with study volunteers, researchers meet the moral goal of maintaining human dignity. That goal may also characterize interactions with IRB staff and committee members.

## *2 Advancing Productive Interactions*

An overarching goal of both researchers and IRBs is the production of ethically responsible research that contributes useful knowledge about humans and their societies (Citro et al., 2003). At the abstract level, their respective responsibilities are complementary. IRBs should assist researchers in meeting ethical standards for the treatment of human research participants. By doing so, IRBs ensure public accountability (Adair, 2001), researchers uphold the ethical codes of their disciplines, and volunteers are protected and are likely to continue to trust in the scientific endeavor. Nevertheless, at extremes, researchers condemn IRBs as obstructionists, and IRBs perceive investigators as dismissive of their real potential to harm study participants. Instead of stressing the complementarity of their responsibilities to create a cooperative relationship, researchers and IRBs often find themselves embroiled in an adversarial relationship. In addition to the call for mutual education noted previously (Citro et al., 2003; DeVries et al., 2004; DuBois, 2004; Oakes, 2002), facilitating the extent to which the ethical principles of respect and justice pervade researcher–IRB interactions may attenuate the negative tenor of the relationship and serve to maintain research integrity (Martinson, Anderson, Crain, & DeVries, 2006).

Mutual education provides a basis for learning about the goals, roles, and behaviors of each group. Beyond gaining information, such education enhances the likelihood that researchers will be likely to imagine what IRBs want, and IRBs may be able to more clearly understand researchers' goals and frustrations. The symbolic interactionist concept of role-taking (Mead, 1934) captures this idea of imagining the interests and behaviors of others, which may prevent misunderstandings and facilitate interaction. Likewise, Eckenwiler (2001)

emphasizes how incorporating the perspectives of others into decision making enhances moral thinking, and DuBois (2004) notes that anticipating the consequences of an action for others and taking them into consideration provides a sounder basis for a moral judgment. Although such processes are fundamental to determining the potential harms or benefits in an empirical study, they apply similarly to directions for (and assessments of) interactions between researchers and the IRB. DuBois (2004) also implies that by adopting a role-taking perspective, researchers are more likely to internalize the norms underlying regulatory demands. Generally, to the extent that disciplines nurture cultures of research ethics independent of IRBs, future generations of scholars may feel less overwhelmed and frustrated by the IRB process (Adair, 2001).

The inclination toward such role-taking may emerge from the communication and interactions that routinely occur between investigators and IRB staff and committee members. Oakes (2002) emphasizes how important it is for researchers to feel free to ask the IRB questions—about IRB proposals and procedures as well as federal regulations. Citro et al. (2003) contend that miscommunication between IRBs and researchers is a source of frustration for both sides. Those authors explicitly reinforce the notion that mutual education will create a better understanding of the functions of the IRB and the concerns of researchers that will pave the way to open, clear communications. At a minimum, all communications should be civil as a means to convey the courtesy that a general principle of respect entails (Sieber, 1992).

Drawing from procedural and interactional justice (Bies, 2001; Tyler & Lind, 1992), respect ensures that individuals in general feel fairly treated. Demonstrating respect involves treating people with sincerity, politeness, and dignity while refraining from deliberately being rude or attacking them. Accepting or esteeming other people's rights is also a way to show respect. Research indicates that authorities who demonstrate respect toward subordinates are likely to be perceived as more legitimate and to elicit greater compliance with their mandates (see Tyler, 2006). In addition, individuals are more likely to view decisions as fair if justifications are provided (e.g., Bies & Shapiro, 1988).

These patterns of findings provide a basis for understanding faulty and promising IRB–researcher interactions. To the extent that “IRB demands are perceived as unjust, researchers react by ignoring IRB demands and bypassing IRBs—and feel justified rather than unethical” (DuBois, 2004, p. 389). Indeed, Martinson et al. (2006) provide evidence that perceived procedural injustice in the contexts in which researchers work enhances the likelihood of scientific misconduct. To be sure, interactions with the IRB are just one element of that context, but such findings reinforce the need for IRBs and researchers alike to act respectfully and fairly, just as the federal regulations mandate with regard to the treatment of study participants. Maintenance of the human dignity of individual actors in their roles as researchers, IRB staff, or IRB members may enhance cooperation that ultimately facilitates the creation of IRB proposals, their review, and compliance with federal regulations.

## V CONCLUSIONS

The production of knowledge is a venerable pursuit. It is, however, not an untethered one. Although multiple definitions of what constitutes ethical research exist, at a minimum, producers of knowledge must be mindful of the implications of their study procedures for the rights and welfare of the people who volunteer for their investigations.

This chapter has outlined ethical concerns regarding social science research in general as well as issues more specific to experimental methodology. Because the outside world intrudes upon research in terms of federal regulations enacted for the protection of human research participants, the chapter has also attempted to reinforce the ethics underlying the regulations and to provide advice with regard to navigating the review process. Social scientists could provide advice to IRBs by bringing their substantive knowledge of power, trust, and organizational justice and methodological expertise to bear upon determining the best ways to ensure the protections of study volunteers (Citro et al., 2003; Sieber, 2004).

Other chapters in this volume provide more explicit and, perhaps, concrete advice with regard to the purpose and construction of laboratory experiments in the social sciences. Such advice is critical because poorly defined studies cannot yield meaningful results and thus cannot justify the demands on study participants (Rosenthal, 1994). In other words, invalid research is also unethical (Sieber, 1992). It is within the purview of IRBs to decline approval of such studies. In contemplating the design of valid studies, ethical directives may constitute methodological challenges (Rosnow, 1997) that, when successfully met, have implications beyond the particular substantive issue under study. They protect human dignity and reinforce trust in the scientific endeavor.

## ACKNOWLEDGMENTS

I thank Rebecca Rousselle and the editors of this volume for their helpful comments. Also, I express my appreciation to Joan Sieber for augmenting my understanding of and dedication to human research protections. Please direct comments to the author at the Department of Sociology, Emory University, Atlanta, GA 30322 or [khegve@emory.edu](mailto:khegve@emory.edu).

## REFERENCES

- Adair, J. G. (2001). Ethics of psychological research: New policies; continuing issues; new concerns. *Canadian Psychology*, 42, 25–37.
- Altman, E. & Heron, P. (Eds.). (1997). *Research misconduct: Issues, implications, and strategies*. Greenwich, CT: Ablex.
- Babbie, E. (2013). *The practice of social research*. Belmont, CA: Wadsworth.
- Baumrind, D. (1985). Research using intentional deception. *American Psychologist*, 40, 165–174.
- Beauchamp, T. L., & Childress, J. F. (2008). *Principles of biomedical ethics*. New York: Oxford University Press.

- Bies, R. J. (2001). Interactional (in)justice: The sacred and the profane. In J. Greenberg & R. Cropanzano (Eds.), *Advances in organizational justice* (pp. 85–108). Stanford, CA: Stanford University Press.
- Bies, R. J., & Shapiro, D. L. (1988). Voice and justification: Their influence on procedural fairness. *Academy of Management Journal, 31*, 676–685.
- Bonetti, S. (1998). Experimental economics and deception. *Journal of Economic Psychology, 19*, 377–395.
- Bordens, K. S., & Abbott, B. B. (2013). *Research and design methods: A process approach*. Boston: McGraw-Hill.
- Boynton, M. H., Portnoy, D. B., & Johnson, B. T. (2012). Exploring the ethics and psychological impact of deception in psychological research. *IRB: Ethics & Human Research, 34*(2), 7–13.
- Christensen, L. (1988). The negative subject: Myth, reality or a prior experimental experience effect? *Personality and Social Psychology Bulletin, 14*, 664–675.
- Citro, C. F., Ilgen, D. R., & Marrett, C. B. (2003). *Protecting participants and facilitating social and behavioral sciences research*. Washington, DC: National Academy Press.
- Cook, K., & Yamagishi, T. (2008). A defense of deception on scientific grounds. *Social Psychology Quarterly, 71*, 215–221.
- Davis, D. D., & Holt, C. A. (1993). *Experimental economics*. Princeton, NJ: Princeton University Press.
- Decker, S. E., Naugle, A. E., Carter-Visscher, R., Bell, K., & Seifert, A. (2011). Ethical issues in research on sensitive topics: Participants' experiences of distress and benefit. *Journal of Empirical Research on Human Research Ethics, 6*(3), 55–64.
- DeVries, R., DeBruin, D. A., & Goodgame, A. (2004). Ethics review of social, behavioral, and economic research: Where should we go from here? *Ethics and Behavior, 14*, 351–368.
- Diviak, K. R., Curry, S. J., Emery, S. L., & Mermelstein, R. J. (2004). Human participants challenges in youth tobacco cessation research: Researchers' perspectives. *Ethics and Behavior, 14*, 321–334.
- DuBois, J. M. (2004). Is compliance a professional virtue of researchers? Reflections on promoting the responsible conduct of research. *Ethics and Behavior, 14*, 383–395.
- Dunn, C. M., & Chadwick, G. (2004). *Protecting study volunteers in research*. Boston: Centerwatch.
- Eckenwiler, L. (2001). Moral reasoning and the review of research involving human subjects. *Kennedy Institute of Ethics Journal, 11*, 37–69.
- Elms, A. C. (1982). Keeping deception honest: Justifying conditions for social scientific research stratagems. In T. L. Beauchamp, R. R. Faden, R. J. Wallace, Jr., & L. Walters (Eds.), *Ethical issues in social science research* (pp. 232–245). Baltimore, MD: Johns Hopkins University Press.
- Epley, N., & Huff, C. (1998). Suspicion, affective response, and educational benefit as a result of deception in psychology research. *Personality and Social Psychology Bulletin, 24*, 759–768.
- Fisher, C. B., & Fyrberg, D. (1994). Participant partners: College students weigh the costs and benefits of deceptive research. *American Psychologist, 49*, 417–427.
- Flagel, D. C., Best, L. A., & Hunter, A. C. (2007). Perceptions of stress among students participating in psychology research: A Canadian survey. *Journal of Empirical Research on Human Research Ethics, 2*(3), 61–67.
- Geller, D. (1982). Alternatives to deception: Why, what, and how? In J. E. Sieber (Ed.), *The ethics of social research: Surveys and experiments* (pp. 38–55). New York: Springer-Verlag.
- Gillespie, J. F. (1999). The why, what, how, and when of effective faculty use of institutional review boards. In G. Chastain & R. E. Landrum (Eds.), *Protecting human subjects* (pp. 157–177). Washington, DC: American Psychological Association.
- Hertwig, R., & Ortmann, A. (2001). Experimental practices in economics: A methodological challenge for psychologists? *Behavioral and Brain Sciences, 24*, 383–451.

- Hertwig, R., & Ortmann, A. (2008). Deceptions in social psychological experiments: Two misconceptions and a research agenda. *Social Psychology Quarterly*, 71, 222–227.
- Hoeyer, K., Dahlager, L., & Lynöe, N. (2005). Conflicting notions of research ethics: The mutually challenging traditions of social scientists and medical researchers. *Social Science and Medicine*, 61, 1741–1749.
- Holmes, D. (1976a). Debriefing after psychological experiments: I. Effectiveness of postdeception dehoaxing. *American Psychologist*, 32, 858–867.
- Holmes, D. (1976b). Debriefing after psychological experiments: II. Effectiveness of postexperimental desensitizing. *American Psychologist*, 32, 868–875.
- Israel, M., & Hay, I. (2006). *Research ethics for social scientists: Between ethical conduct and regulatory compliance*. Newbury Park, CA: Sage.
- Kelman, H. C. (1982). Ethical issues in different social science methods. In T. L. Beauchamp, R. R. Faden, R. J. Wallace, Jr., & L. Walters (Eds.), *Ethical issues in social science research* (pp. 40–97). Baltimore, MD: Johns Hopkins University Press.
- Korn, J. H. (1997). *Illusions of reality: A history of deception in social psychology*. Albany, NY: University of New York Press.
- Landrum, R. E., & Chastain, G. (1999). Subject pool policies in undergraduate-only departments: Results from a nationwide survey. In G. Chastain, & R. E. Landrum (Eds.), *Protecting human subjects* (pp. 24–42). Washington, DC: American Psychological Association.
- Martin, M. W., & Sell, J. (1979). The role of the experiment in the social sciences. *The Sociological Quarterly*, 20, 581–590.
- Martinson, B. C., Anderson, M. S., Crain, A. L., & DeVries, R. (2006). Scientists' perception of organizational justice and self-reported misbehaviors. *Journal of Empirical Research on Human Research Ethics*, 1(1), 51–66.
- McBurney, D. H., & White, T. L. (2012). *Research methods*. Belmont, CA: Wadsworth/Thomson.
- Mead, G. H. (1934). *Mind, self, and society*. Chicago: University of Chicago Press.
- Milgram, S. (1974). *Obedience to authority: An experimental view*. New York: Harper & Row.
- Moyer, A., & Franklin, N. (2011). Strengthening the educational value of undergraduate participation in research as part of a psychology department subject pool. *Journal of Empirical Research on Human Research Ethics*, 6(1), 75–82.
- National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. (1979). *The Belmont report: Ethical principles and guidelines for the protection of human subjects* (Federal Register Document No. 79-12065). Washington, DC: U.S. Government Printing Office.
- Neuman, W. L. (2011). *Basics of social research: Qualitative and quantitative approaches*. Boston: Pearson.
- Oakes, J. M. (2002). Risks and wrongs in social science research: An evaluator's guide to the IRB. *Evaluation Review*, 36, 443–479.
- Oczak, M., & Niedźwieńska, A. (2007). Debriefing in deceptive research: A proposed new procedure. *Journal of Empirical Research on Human Research Ethics*, 2(3), 49–59.
- Ortmann, A., & Hertwig, R. (2002). The costs of deception: Evidence from psychology. *Experimental Economics*, 5, 111–131.
- Pascual-Leone, A., Singh, T., & Scoboria, A. (2010). Using deception ethically: Practical research guidelines for researchers and reviewers. *Canadian Psychology*, 51, 241–248.
- Ripley, E. B. (2006). A review of paying research participants: It's time to move beyond the ethical debate. *Journal of Empirical Research on Human Research Ethics*, 1(4), 9–20.
- Rosenthal, R. (1994). Science and ethics in conducting, analyzing, and reporting psychological research. *Psychological Science*, 5, 127–134.

- Rosnow, R. L. (1997). Hedgehogs, foxes, and the evolving social contract in psychological science: Ethical challenges and methodological opportunities. *Psychological Methods*, 2, 345–356.
- Sell, J. (2008). Introduction to deception debate. *Social Psychology Quarterly*, 71, 213–214.
- Sieber, J. (1992). *Planning ethically responsible research: A guide for students and internal review boards*. Newbury Park, CA: Sage.
- Sieber, J. (1999). What makes a subject pool (un)ethical? In G. Chastain & R. E. Landrum (Eds.), *Protecting human subjects* (pp. 43–64). Washington, DC: American Psychological Association.
- Sieber, J. (2003). *Risk, benefit, and safety in human research*. Unpublished manuscript, Hayward: California State University.
- Sieber, J. (2004). Empirical research on research ethics. *Ethics and Behavior*, 14, 397–412.
- Sieber, J., Iannuzzo, R., & Rodriguez, B. (1995). Deception methods in psychology: Have they changed in 23 years? *Ethics and Behavior*, 5, 67–85.
- Sieber, J., & Saks, M. J. (1989). A census of subject pool characteristics and policies. *American Psychologist*, 44, 1053–1061.
- Singer, E., & Couper, M. P. (2008). Do incentives exert undue influence on survey participation? Experimental evidence. *Journal of Empirical Research on Human Research Ethics*, 3(3), 49–56.
- Singer, E., & Levine, F. J. (2003). Protection of human subjects of research: Recent developments and future prospects for the social sciences. *Public Opinion Quarterly*, 67, 148–164.
- Tyler, T. R. (2006). Psychological perspectives on legitimacy and legitimization. *Annual Review of Psychology*, 57, 375–400.
- Tyler, T. R., & Lind, E. A. (1992). A relational model of authority in groups. *Advances in Experimental Social Psychology*, 25, 115–191.
- Veatch, R. M. (1987). *The patient as partner*. Bloomington, IN: Indiana University Press.
- Waite, B. M., & Bowman, L. L. (1999). Research participation among general psychology students at a metropolitan comprehensive public university. In G. Chastain, & R. E. Landrum (Eds.), *Protecting human subjects* (pp. 69–85). Washington, DC: American Psychological Association.
- Zelditch, M., Jr. (1969). Can you really study an army in the laboratory? In A. Etzioni (Ed.), *Complex organizations* (pp. 528–539). New York: Holt, Rinehart & Winston.
- Zimbardo, P. G. (1973). On the ethics of intervention in human psychological research: With special reference to the Stanford prison experiment. *Cognition*, 2, 243–256.
- Zimbardo, P. G., Banks, W. C., Haney, C., & Jaffe, D. (1973, April 8). The mind is a formidable jailer: A Pirandellian prison. *The New York Times Magazine, Section*, 6, 38–60.

## Chapter 3

# Logical and Philosophical Foundations of Experimental Research in the Social Sciences

Shane R. Thye

*University of South Carolina, Columbia, South Carolina*

## I INTRODUCTION

Social scientists of various academic stripes depict the experimental method in diverse ways—sometimes good, mostly bad. Lieberson (1985, p. 228) asserts that “the experimental simulation fails at present in social science research because we are continuously making counterfactual conditional statements that have outrageously weak grounds.” Psychologists, who normally embrace experimental methods, have made even more radical claims: “The dissimilarity between the life situation and the laboratory situation is so marked that the laboratory experiment really tells us *nothing*” (Harré & Secord, 1972, p. 51, italics in original). Even writers of popular methods textbooks make critical observations. Babbie (1989, p. 232) argues that “the greatest weakness of laboratory experiments lies in their artificiality.” There are numerous costs and benefits associated with any research method, but for whatever reason, the experimental method seems to inspire its fair share of critics.

If you are reading this, you are probably a student who has been assigned this chapter as part of a research methods course, or perhaps you are just interested in the logic and philosophical underpinnings of experimental methods. You may even, intuitively, believe that some or all of the preceding critiques are valid. After all, it has been pointed out for decades that experiments *are* artificial, and indeed they are (Webster & Kervin, 1971). The term *artificial* usually means “novel” in the sense that subjects are placed into situations (e.g., a lab) that are unusual or perhaps unnatural. But when you stop to think about it, don’t all research methods involve somewhat novel, unusual, or artificial conditions? What is so natural about filling out a survey or being probed through an in-depth interview? We do not routinely do these things. The experimental situation is just one kind of situation that people respond to, no more or less novel than

your very first visit to New York City, going on a first date with someone you have met online, or giving birth to your first child. Novelty in and of itself is a part of everyday life and does not really distinguish experimentation from other popular research methods.

Henshel (1980) correctly pointed out more than three decades ago that *unnatural* experimentation is desirable and, in fact, the dominant mode in fields such as medicine, chemistry, and physics. I can do no better than he did, so I will use his example. Consider the modern airplane. We have all seen footage of early attempts at flight using awkward, wing-flapping devices pushed off the edge of cliffs only to plunge to the ground. They never flew. Those inventors attempted to mimic nature—that is, the way that birds obtain lift by flapping their wings—and they all failed. Some even went so far as to imitate the layered structure of a bird's secondary wing feathers on their flying machines. It did not help. It was not until the Wright brothers began to experiment with a fixed-wing aircraft that flight was finally obtained. Nowhere in nature does a fixed-wing aircraft exist, but because of unnatural experimentation, we enjoy the convenience of modern air travel.<sup>1</sup> Some of the most important conveniences in contemporary life are the result of unnatural experimentation, from the X-rays that diagnose disease to the silicone chips that power your smartphone. My hope is that by the end of this chapter, I can show you that the novelty and artificiality of the experimental method are enormous benefits, not liabilities (see also Webster & Kervin (1971) for a statement as to why).

Perhaps the most important feature of controlled experimental research is that it gives one the ability to claim, with some degree of confidence, that two factors are causally linked. Social scientists are in the business of establishing causal laws that explain real-world events. This objective is realized through the painstaking building and systematic testing of scientific theory. As a sidebar, to say that one has an explanation for some phenomenon is to say that there is a well-supported scientific theory of that phenomenon. For example, if you “explain” why skydivers descend to the earth by invoking the notion of gravity, what you really mean is that the observed descent conforms to the theory of gravitational forces. To date, controlled experimentation is the most widely embraced method for establishing scientific theory because it allows scientists to pinpoint cause–effect relations and eliminate alternative explanations. Laboratory experimentation is the gold standard for isolating causation because the logic of experimental research embodies the logic of scientific inquiry. *Thus, the advantage of the experimental method is that it allows one to see the world in terms of causal relations.*

Ironically, this also is a heavy burden carried by the experimental scientist. Consider a headline that read, “Bottled Water Linked to Healthier Babies.” If it is true that expectant mothers who drink bottled water tend to have babies with

---

1. It is true that a soaring bird does not flap its wings, but the appropriate analogy for a soaring bird is an unpowered glider. Neither has the capacity to obtain lift like a powered fixed-wing aircraft.

fewer birth defects, higher birth weights, and other health benefits, this simple yet powerful intervention may carry major health consequences. However, my excitement over the power of water was short-lived as my newfound “parent” identity had to be reconciled with the “experimentalist” in me who also had an opinion. During that transformation, a number of questions emerged. What is the *theoretical connection* between bottled water and healthy babies? Is there a biological mechanism, or could the effect be due to other factors? The latter issue concerns whether or not the relationship is real. So-called spurious factors are unrecognized causes (e.g., socioeconomic status) that produce the illusion that two things (drinking bottled water and having healthy babies) are causally linked. Could it be that socioeconomic status causes both? Perhaps moms who can *afford* bottled water have healthier babies because they can *afford* to get better prenatal care, join gyms, eat healthier food, take vitamins, receive treatment at elite hospitals, and so on. For the same reason, I would lay heavy odds that moms who drive new BMWs also have healthier babies than moms who do not own a car, but that would be unlikely to grab the headlines.

In what follows, I examine the underlying logic of experimental design and analysis. I consider how experimental research bears on establishing of causation, explore recent critiques of Fisherian ([Fisher, 1935, 1956](#)) methods and show how these are flawed, and examine various forms of experimental design. The aim is not to provide a comprehensive discussion of all facets of experimentation. Instead, I hope to illuminate the logic and philosophical underpinnings of experimental research, dispel a number of myths and misconceptions, and generally excite the reader about the prospects of building scientific theory via experimentation. Along the way, I consider a number of issues germane to all research, such as how evidence bears on theory, notions of causation, and the logic of applying or generalizing findings to other settings.

## II CLUES TO CAUSATION

The notion of “causality” has always been a challenge, in part, because causation is not directly observable. Rather, causation must be inferred from some manner of evidence. This section considers the various ways that scientists think about causality and the methods scientists use to infer that two phenomena are causally linked. The concept of causality has a twisting and convoluted history in the philosophy of science, and there are many ongoing discussions and debates. Most notions of causation are traced to [Aristotle \(340 bc/1947\)](#), who offered four different conceptions of causation. Of these, Aristotle’s *efficient cause* captures the notion that one event (X) sets into motion, forces, creates, or makes another event (Y) occur. Although this kind of causation corresponds well with the everyday meaning of the term, it has become the focal point of controversy and debate.

Galileo developed an alternative causal ontology that equated causation with necessary and sufficient conditions. Galileo argued that to say event *a* caused

event *b* was to say that event *a* is a necessary and sufficient condition for event *b* (see Bunge, 1979, p. 33). A *necessary condition* exists when event *b* *never* occurs in the absence of event *a*. That is, event *b* follows event *a* with 100% regularity. A *sufficient condition* exists when event *b* *always* follows event *a* with perfect regularity. To illustrate, a lawnmower will only start if it has the correct kind of fuel in its fuel tank. Thus, the proper kind of fuel is a necessary condition for starting the mower because it will *never* start without it. At the same time, fuel alone is not enough to start the mower; it requires a working engine and all requisite parts. This means that proper fuel is not a sufficient condition to make a mower start. The combination of the proper fuel and a working engine are necessary and sufficient conditions; when both are in place, the mower *always* starts, and if either is missing the mower *never* starts.

The philosopher David Hume (1748/1955) took the more radical position associated with British empiricism. Prior to Hume, popular notions of causation traced to Aristotle involved one event forcing, or setting into motion, another event. Hume offers a softer notion that robs causation of its force. He argued that we can never directly observe a causal force in operation, but instead, all we can observe is the conjunction or correlation of two events that we presume are causally linked. Hume used the term “constant conjunction” to describe the situation in which X always occurs in the presence of Y. Thus, for Hume and the empiricists, causation is an elusive thing—we can never be sure that one thing causes another or that events correlated today will be correlated tomorrow. More radical Humeans, such as Bertrand Russell, reject the idea of causation altogether. In his now infamous 1913 paper, Russell wrote, “The law of causality, I believe, like much that passes muster among philosophers, is a relic of a bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm” (Russell, 1913).

Despite the historical disagreements surrounding notions of causation, the majority of social scientists generally agree on certain basic requirements that must be satisfied to support causal inference (Davis, 1985; Hage & Meeker, 1988). One can think of these requisites as “clues” to assess if a relation is truly causal or not. Next, I consider six conditions that scientists use to assess causation. This discussion focuses not on points of disagreement but, rather, on the general principles on which there is consensus.

## A Covariation

When a cause occurs, then so should its effect; when a cause does not occur, then neither should its effect (this is one of the basic maxims underlying Mill’s canons of inference discussed later). In short, causes and their effects should covary or be correlated. At the same time, it is very important to remember that things which covary may not be causally related. For instance, the price of bourbon is correlated with the price of new cars in any given month: when bourbon is expensive, cars are expensive. In this case, the prices of bourbon and

new cars are not causally linked, but both are caused by the prevailing economic conditions. As Hume would agree, it is easy to focus on the correlation between car prices and bourbon while blindly missing the underlying causal force. As such, this leads to a very important principle: *correlation alone does not imply causation*. Sifting causation from correlation is a focal issue in virtually all social science research.<sup>2</sup>

## B Contiguity

There is always some time lag between a cause and its effect. When the time lag between a cause (a paper cut) and its effect (bleeding) is short, we say the two events are contiguous. Some cause–effect relations are contiguous, whereas others are not (e.g., conception and childbirth). In general, social scientists presume that noncontiguous causes set into motion other processes that, in turn, have effects at a later point in time. For example, greater parental education can lead to a variety of lifestyle benefits (e.g., greater income, more social capital, and advanced reading and verbal skills) that have effects on their children’s educational attainment. Importantly, however, a causal claim between parent and child education levels is not warranted unless there is a theory that specifies the intermediary factors occurring between the two points in time. For example, more educated parents are more likely to read to and with their children than are uneducated parents, and the greater reading by those children helps them succeed in school.

## C Time and Asymmetry

One of the most basic principles of causation is that if X causes Y, then X must occur before Y in time. It is important to note that simply because X precedes Y does not mean X causes Y. Just because it rained before I failed my exam does not mean the rain caused me to fail (if I do make the connection between rain and failure, I am committing the *post hoc fallacy*). A related idea is that causation is assumed to run in one direction, such that causes have asymmetric effects. That is, we cannot simultaneously assert that X causes Y and Y causes X. Also, although the notion of *reciprocal causation*  $X_1 \rightarrow Y \rightarrow X_2$  seems to violate the assumptions of time and asymmetry, it does not because the X at time 2 is not the same X as the X at time 1. Instead, reciprocal causation implies that cause–effect sequences dynamically unfold. For instance, cybernetic feedback systems involve reciprocal causation in which the cause–effect relationships are reversed through time. The “cruise control” feature in your car operates on this

---

2. The basic experimental design provides an elegant solution to the problem of sorting causal relations from correlations. Because this problem has an even tighter stranglehold on those who deploy survey, historical, qualitative, or ethnographic methods, researchers in these domains often use the experiment as a template to design their own studies (Lieberson, 1991).

principle. That is, an initial cause (the deceleration of your car) triggers an effect (an increase in engine RPM) that in turn feeds back on that initial cause (the acceleration of your car).

## D Nonspuriousness

When two things occur together but are truly caused by some third force, the original relationship is said to be *spurious*. There are many unusual spurious relations that illustrate the point. For instance, few people know that there is a strong positive correlation between ice cream sales and rape. That is, in months when ice cream sales are high, many rapes are reported. Does this mean that ice cream sales are somehow causally linked to the occurrence of rapes? The answer is of course not! The relationship is spurious because both factors are caused by a third factor—temperature. In warmer months, more ice cream is sold and there are more sexual predators frequenting outdoor venues and social gatherings. In cold weather, both factors are attenuated. It would be a mistake to believe that rapes and ice cream sales are casually related without seeking alternative explanations. To establish causation, one must be able to, with some degree of confidence, rule out alternative explanations. Ruling out alternatives is a key activity in science, and as demonstrated later, experimental research provides the best known method for doing so.

## E Consistency

Philosophers and scientists have long debated whether one should think of causality as *deterministic* (i.e., that a given cause X will always lead to effect Y) or *probabilistic* (i.e., that the presence of cause X will increase the likelihood of effect Y). Early writers such as [Galileo \(1636/1954\)](#) and [Mill \(1872/1973\)](#) leaned toward the deterministic end of the spectrum. However, with the advent of modern statistical tools, the majority of contemporary philosophers and social scientists evoke probabilistic notions of causality. To illustrate, there is abundant evidence that smoking causes a higher rate of lung cancer. Still, not *every* smoker develops lung cancer, and not *every* lung cancer victim is a smoker. It seems more accurate to say that smoking increases the probability of lung cancer. [Hage and Meeker \(1988\)](#) argue that probabilistic notions of cause are preferable for three reasons. First, there can be unrecognized countervailing causal forces (e.g., antibodies or gene combinations that make people resilient to cancer) that play a role. Second, most phenomena are affected by a multiplicity of causes (e.g., body chemistry or exposure to carcinogens) that can interact to obfuscate true causal relations. Third, chaos and complexity theories have shown that both natural and social phenomena can behave in unpredictable and nonlinear ways ([Waldrop, 1992](#)). To illustrate, those with advanced cancer can for some inexplicable reason go into remission. The upshot is that the deterministic views of causation may be overly simplistic.

## F Theoretical Plausibility

Finally, scientists always view causal claims about the world with one eye trained on established scientific laws. When claims about the world violate or are inconsistent with those laws, then without unequivocal evidence to the contrary, the confirmation status of those claims is questionable. Simply stated, extraordinary claims require extraordinary evidence. For instance, Newton's law of inertia states that unless acted upon, a body at rest stays at rest and a body in motion stays in motion. Based on this thinking, many have attempted to build "perpetual motion" machines ignoring the broader context of other physical laws. A true perpetual motion machine (if one could be built) would, in fact, violate several existing scientific laws. For example, such a machine would need to consume no energy and run with perfect efficiency (thus violating the second law of thermodynamics) or produce energy without consuming energy as it runs (thus violating the first law of thermodynamics). Thus, although Newton's law of inertia *suggests* that perpetual motion machines are possible, in the context of the laws of thermodynamics, it seems that such machines are very unlikely to ever be produced. It should not be surprising that, to date, dreams for perpetual motion machines have remained just that.

## III MILL'S CANONS AND INFERRING CAUSALITY

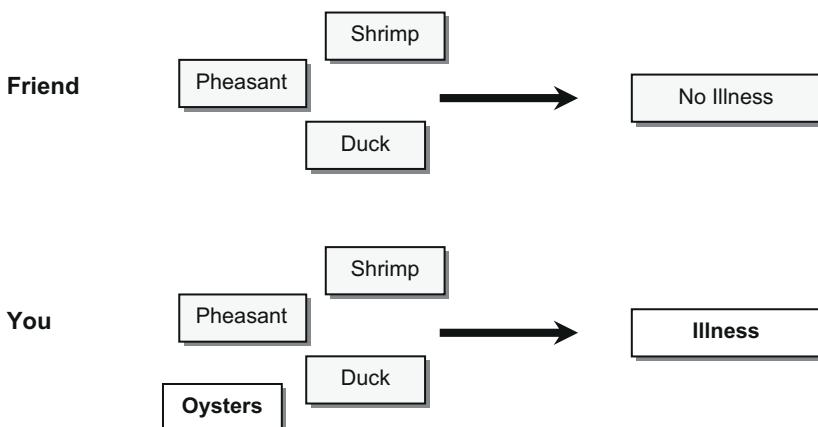
What is the logical connection between data and causation? Asked differently, how does one infer causation based on the outcome of some empirical test? Many, if not all, social scientists operate by approximating a model of evidence that can be traced to [John Stuart Mill \(1872/1973\)](#), who was influenced by Sir Francis Bacon and subsequently influenced [Sir Ronald A. Fisher \(1935, 1956\)](#). Mill presumed that nature is uniform, and as such, if a cause–effect relationship occurred once, he presumed it would occur again in similar circumstances. Mill developed five methods (or canons) to assess causation; these canons lie at the heart of contemporary inference and experimental design today. Here, I briefly discuss two of the more important canons: the method of difference and the method of agreement. Whereas the method of difference aims to find a lone *difference* across two circumstances, the method of agreement seeks to find a lone similarity.

*The method of difference:* "If an instance in which the phenomenon under investigation occurs, and an instance in which it does not occur, have every circumstance in common save one, that one occurring only in the former; the circumstance in which alone the two instances differ, is the effect, or the cause, or an indisputable part of the cause, of the phenomenon" (1872/1973, p. 391).

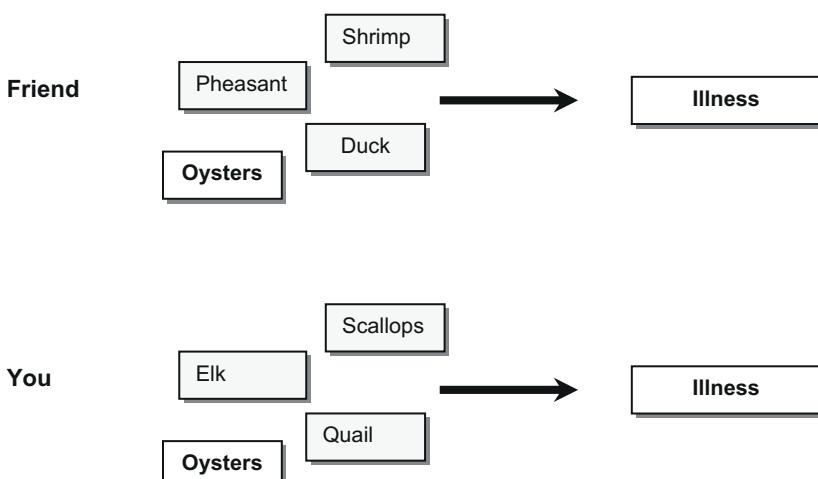
*The method of agreement:* "If two or more instances of the phenomenon under investigation have only one circumstance in common, the circumstance in which alone all the instances agree, is the cause (or effect) of the given phenomenon" (1872/1973, p. 390).

Perhaps the best way to illustrate the methods is by example. Imagine that you and a friend attend a wild game cookout, sample a variety dishes, and later that night you feel ill but your friend does not. How would you determine the cause of the illness? Imagine that you and your friend both sampled shrimp, pheasant, and duck, while you also enjoyed oysters but your friend did not. The top panel of [Figure 3.1](#) illustrates this scenario. Notice that there are two differences between you and your friend: you dined on oysters and then became ill

### Panel A The Method of Difference



### Panel B The Method of Agreement



**FIGURE 3.1** Mill's method of difference and method of agreement.

while your friend did neither. Mill's method of difference suggests that oysters are the cause of the illness because it is the single condition that distinguishes illness from health. The same method can also be used as a method of elimination. Notice that both you and your friend ate shrimp, pheasant, and duck, but only you got ill. Thus, shrimp, pheasant, and duck can be eliminated from the list of possible causes. In this way, the method can be used to eliminate alternative causes because *a deterministic cause that remains constant can never produce effects that are different*.

Now consider the method of agreement, which indicates that if a single factor (eating oysters) occurs in conjunction with a common effect, then that single factor is the likely cause. Assume that your friend enjoyed shrimp, pheasant, duck, and oysters, while you ate scallops, elk, quail, and oysters. Later that evening, you both became ill. Because eating oysters is the common denominator preceding illness, you might infer that oysters are the cause. This illustrates another important principle: *a causal factor that differs across two circumstances can never generate precisely the same effect*. In this case, if eating shrimp were the true cause of illness, because your friend ate shrimp and you did not, we would expect your friend (but not you) to be ill. The method of agreement once again suggests that the oysters are suspect.

## A Limitations of Mill's Canons

Mill's (1872/1973) method of difference and method of agreement provide useful guidelines for thinking about causation. Even so, it is now recognized that the canons are limited in a number of regards. A good critique of the methods can be found in Cohen and Nagel (1934; see also B. Cohen, 1989), and it may be useful to summarize their analyses here. Cohen and Nagel (1934, p. 249) point out that Mill's methods are neither "methods of proof" nor "methods of discovery" (see also Hempel, 1965). In terms of proof, the canons generally presume that *all other possible causes* are contained in the factors examined. In the previous example, we presume that all possible causes of illness reside in the food that was eaten. However, the methods cannot *prove* this to be definitively so because there *could* be unrecognized causes in things besides the food. Imagine that you and your friend have an undetected mild shellfish allergy that only reacts when two or more varieties of shellfish are eaten simultaneously. Referring back to Figure 3.1, in every case of illness, both oysters and one other shellfish (shrimp or scallops) were consumed. As such, it might not be oysters per se causing the illness but, rather, the unique combination of food and the allergy—what statisticians call an *interaction effect*. The method of difference and method of agreement are blind to this possibility.

Second, Mill's (1872/1973) methods cannot be used as "methods of discovery" because they presuppose that one can identify, *a priori*, potential causes of the illness. In Figure 3.1, differences and similarities across food items are analyzed as potential causes for the illness. However, there are other potential

causes that are unmeasured or unknown. This includes personal factors (food consumed before the party), social factors (contact with other sick people), historical factors (flu or allergy season), or even genetic or biological factors (a weak immune system) that might fuel illness in a difficult to detect way. In principle, there are an *infinite* number of additional causes that could contribute to illness, and the lion's share of these will be unmeasured and unknown. Again, the methods are not methods of discovery because they are oblivious to possible alternatives.

In summary, Mill's canons suffer two limitations: *proving* that a single factor is the unique cause and *discovering* causes beyond the immediate situation. Although it may not be immediately obvious, both limitations stem from a single issue: *any phenomenon may have an infinite number of intertwined causes that researchers cannot measure or identify*. Given this seemingly insurmountable problem, one might guess that Mill's (1872/1973) methods would have withered on the vine. Indeed, that might have occurred had it not been for the statistician R. A. Fisher,<sup>3</sup> who rescued the methods. Fisher (1935, 1956) did that by providing a logical and methodological basis for reducing or altogether eliminating the troublesome set of alternative explanations. Next, I illustrate Fisher's solution and other central features of experimental research through a detailed example. Following this, I reconsider Mill's (1872/1973) methods in view of Fisher's solution and Cohen and Nagel's (1934) critique.

## IV FISHER'S SOLUTION AND HALLMARKS OF EXPERIMENTATION

The confluence of three features makes experimental research unique in scientific inquiry: random assignment, manipulation, and controlled measurement. To illustrate, presume that a researcher is intrigued by the relationship between viewing television violence and childhood aggression, a topic that captures much scrutiny. A typical hypothesis is that viewing television violence increases subsequent aggression by imitation. Correspondingly, let us assume that a researcher plans to study 100 fourth-grade children, their viewing habits, and their aggressive behaviors. The overall experimental strategy involves three distinct phases. First, the researcher must *randomly assign* each child to one of two groups. The first group will be exposed to violent television (called the treatment group) and the second group will be exposed to nonviolent television (called the control group). Second, the researcher *manipulates* the content of the violent television such that one program contains violence while the other does not. Third, the researcher *measures* aggressive behavior across the groups in a controlled environment. Let us consider each feature in more detail.

---

3. Despite the tremendous vision Fisher possessed as a statistical prodigy and brilliant geneticist, he was, ironically, severely myopic. He claimed that this aided, rather than hindered, his creative insight because he was forced to rely more heavily on mental instead of physical representations.

## A Random Assignment

Fisher (1935, 1956) offered the concept of random assignment as a way to hold constant spurious causes. *Random assignment* is defined as the placement of objects (people or things) into the conditions of an experiment such that each object has exactly the same probability of being exposed to each condition. Thus, in the two condition experiment described previously, each child has exactly a 50% chance of being assigned to the experimental or control group. This procedure ensures that each group contains approximately 50 children, half male and half female, to within the limits of random chance. The benefit of random assignment is that it *equates at the group level*. In other words, random assignment ensures that the two groups are equal in terms of all *historical factors* (e.g., divorced parents, childhood poverty, and being abused), *genetic factors* (eye color, chromosome distribution, blood type, etc.), *physical factors* (gender, height, weight, strength, etc.), *personality factors* (preferences, values, phobias, etc.), and *social factors* (role identities, dog ownership, etc.). The reason is that each factor has exactly the same probability of appearing in each group. The brilliance of random assignment is that it mathematically equates the groups on factors that are known, unknown, measured, or unmeasured.<sup>4</sup>

A common criticism is that experimental results are biased if the traits under the control of random assignment interact with the dependent variable. This idea, however, is slightly off the mark. For instance, assume that: (1) both males and females respond aggressively, but differently, to watching violent television; and (2) our measure of aggression is sensitive to male, but not female, aggression. The result of the experiment would correctly show that males in the treatment group respond more aggressively than males in the control group *on this measure of aggression*. It would also correctly show no difference between the treatment and control females *on this measure of aggression*. Strictly speaking, the problem here is not one of random assignment or even one of incorrect inference. Instead, the problem is that one may not have a robust and valid measure of hostility that captures the kind of aggression expected to occur in both males and females. The relationship between exposure to violent television and aggressive behavior is properly guided by the theory that links the two phenomena. Such an interaction suggests a problem with the theory or a problem with the measurement procedures, but not the experiment per se.

## B Manipulation

Following random assignment, the independent variable is manipulated such that the treatment group is exposed to violent images while the control group is not.

---

4. There are deep-rooted statistical reasons to employ random assignment. For instance, violating random assignment can cause observations to be correlated, which can bias any ensuing statistical test, inflate the standard error of that test, or both. Thus, random assignment rests on logical and statistical foundations.

Ideally, the manipulation would be exactly the same in both conditions except for the factor hypothesized as causal. In our example, the researcher could ask the experimental group to watch a television program of a couple engaged in a financial dispute that ends violently. The control group could watch the exact same couple end the exact same dispute in a nonviolent manner. Notice that in the context of random assignment, the basic experiment is comparable to Mill's method of difference in that the two groups are equated on virtually all factors except the independent variable. Apodictically, the Fisherian principle of random assignment is the cornerstone of experimental research, and in conjunction with controlled manipulation, these procedures render the method of difference workable.

## C Controlled Measurement

The final step is to measure the dependent variable (aggressive behavior) in a controlled environment. Ideally, individuals from the treatment and control group would have the opportunity to aggress in exactly the same manner toward the same target. In actual research, aggression has been measured in a variety of ways, including the delivery of electric shock, hitting a doll with a mallet, or the slamming down of a telephone. Importantly, any difference between the treatment and control groups can only be attributed to the independent variable because, in principle, this is the only factor on which the two groups differ (remember that a cause that does not vary can have no effect). Overall, the experimental method is the method of causal inference because it equates two or more groups on all factors, manipulates a single presumed cause, and systematically records the unique effect.

How does the method measure up with respect to providing information on causation? Recall that causal inferences require information on three empirical cues (covariation, contiguity, and temporal ordering) and three other criteria (nonspuriousness, consistency, and theoretical plausibility). Overall, the basic experiment does an excellent job of attending to these matters. Evidence for covariance comes in the form of the effect appearing in the experimental group but not the control group. Also, the researcher has information on contiguity because he or she controls the time lag between the factors under investigation. Furthermore, because the experimenter manipulates the cause before the effect, the temporal ordering is correctly instated. Of course, random assignment controls for potentially spurious factors. Finally, the experiment is a repeatable event, and its data are always considered in the context of established theory.

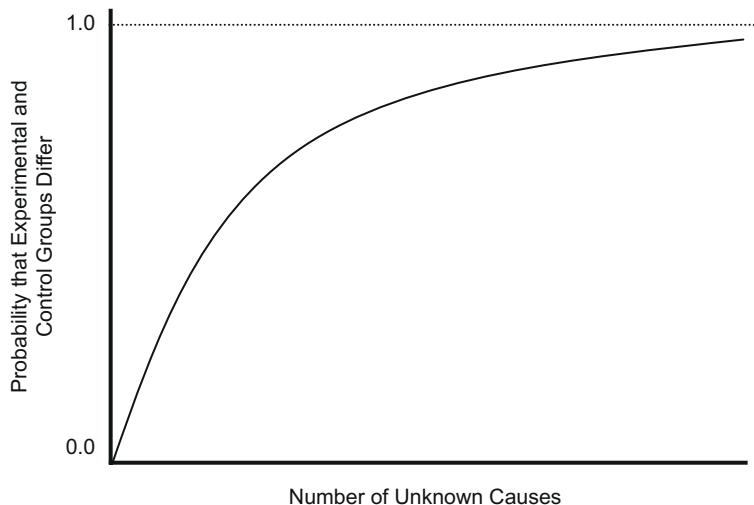
## V FISHER'S PREMATURE BURIAL AND POSTHUMOUS RESURRECTION

The basic experiment approximates Mill's (1872/1973) method of difference in structure and design. However, recall that Mill's method was deemed intractable

because there could be an infinite number of unknown and unknowable causes adding to (or interacting with) the presumed cause. A related problem is that Mill invokes a deterministic view of causation wherein empirical outcomes occur with perfect regularity, which of course never occurs (Willer & Walker, 2007). Fisher (1935) developed the principle of random assignment to remedy the ailing method. When subjects are randomly assigned to conditions, any unknown and unknowable factors are distributed equally and thus: (1) *the causal effects of those unknown and unknowable factors will be equated across experimental and control groups*; and (2) *factors that are the same can never cause a difference*. Thus, any differences can be reasonably attributed to variation from the independent variable. Although Fisher's principle of random assignment is widely recognized as the panacea, others still believe the method is grievously ill.

Bernard Cohen (1989, 1997) asserts that random assignment does not remedy the problem. He argues that as the set of unknown or unknowable causes grows large, the probability that the experimental and control groups will differ on at least one of these causal factors approaches 1.0. This relationship is shown in Figure 3.2. Furthermore, Cohen notes that random assignment only equates experimental and control groups with infinite sample sizes, and of course, real experiments always employ finite samples. He therefore concludes that there is always some unknown causal factor operating in any experiment, and as such, the original problems that plagued Mill's (1872/1973) methods remain unresolved. In the end, Cohen (1997) does the only befitting thing and offers "a decent burial" for J. S. Mill and R. A. Fisher.

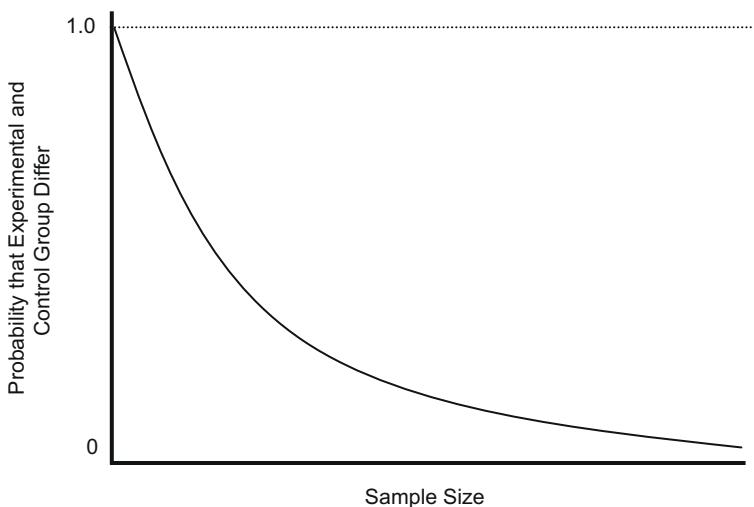
At first blush, Cohen (1989, 1997) appears to have brought the illness out of remission. However, he may have been premature in laying our progenitors



**FIGURE 3.2** The relationship between unknown causes and probability that experimental and control groups differ.

and their methods to rest. His analysis centers on: (1) the number of potential alternative causes; and (2) the number of subjects in the experimental and control groups—claiming that the former is too large and the latter is too small for random assignment to properly operate. However, in a mathematical sense, [Cohen \(1997\)](#) overestimates both the number of alternative causes that *do* exist and the sample size that *is* required for random assignment to work properly. In terms of sample size, [Cohen \(1997\)](#) ignores the straightforward statistical relation between sample size and the nature of alternative causes. That relationship is described by the *law of large numbers*. This law dictates that as the size of the experimental and control groups increases, the average value of any factor differentiating those groups approaches a common value (i.e., the population value for that factor). This law also has implications for the *number* of alternative causes. The law implies that as group size increases and alternative causal factors become equated, the number of *differentiating possible causes will diminish at an exponential rate*. As this occurs, the probability that the experimental group differs from the control group, on any single factor, approaches zero.<sup>5</sup> The relationship between sample size and probability of differentiating factors is shown in [Figure 3.3](#). Thus, probability theory and the law of large numbers mitigates (but does not totally eliminate) the issue of unknown factors.

There are additional logical and empirical grounds that further salvage the method. Logically, it is important to distinguish factors that *could make*



**FIGURE 3.3** The relationship between sample size and probability that experimental and control groups differ.

5. For instance, with just 23 randomly selected people, there is a 50% chance 2 of them will share a birthday ([Paulos, 1988](#))!

a difference from factors that *do make a difference*. Not all of the factors that make experimental groups different from control groups are relevant to the dependent variable, and therefore, not all factors must necessarily be equated. Many differences simply do not matter. For instance, physical theories do not consider the color of falling objects for the same reason that bargaining theories do not consider the height of the negotiator; neither factor is relevant to the theoretical processes and phenomena of concern (Willer & Walker, 2007). Scientists specifically control for *theoretically relevant* factors in a given study, and in true experiments they also control for *possibly relevant* factors using random assignment. Cohen heavily emphasizes those factors that are possibly relevant. However, given that laboratory research selectively focuses on specific theoretical problems, most of the infinite set of possibly relevant differences probably will be irrelevant.

Now, let's play devil's advocate and presume Cohen (1997) is correct in his assertion. That is, assume that a very large number of factors differentiate experimental and control groups and that these factors are relevant to the phenomena under consideration. From a statistical standpoint, ironically, even this situation does not pose a problem for the internal validity of the experiment. The *central limit theorem* explains how those factors would be distributed. This theorem suggests that for a very large number of independent causal factors differentiating experimental and control groups, the aggregate effect of those causal factors would quickly converge and become normally distributed. The shape of the distribution is important. Some causal factors would have positive effects on the dependent variable; others would have negative effects. When subjects are randomly assigned to conditions, the overall impact of these factors is to add *random variance* (or noise) to the dependent variable, and that noise would be normally distributed with an expected value of 0. Thus, *random assignment and the central limit theorem ensure the errors are normally distributed and guarantee there is no overall impact of these variables*. This means that even if the identified problems are real, they do not affect the basic logic of experimental inference or impact the validity of the method.

In summary, the method of difference and the principle of random assignment provide the logical and statistical foundation for contemporary experimental design and analysis. Mill and Fisher left in their wake a powerful set of analytic and statistical tools that have become the method of understanding causality in science. I have shown that the problems identified by Cohen (1989, 1997) and others are not problematic in a statistical or pragmatic sense. In practice, if the experimental method were flawed in the ways detailed previously, we would never expect to find consistent results produced by experimental research, nor would we expect to find cumulative theory growth in areas so informed. However, even Cohen (1989) acknowledges that cumulative research programs informed by experimental work abound. Despite the fact that both Mill and Fisher have long since passed away, the legacy of experimental testing they left behind continues to flourish.

## VI SIMPLE DESIGNS AND THREATS TO INTERNAL VALIDITY

At this juncture, it seems that the logic of experimentation provides a fairly bulletproof way to establish causation. It would be a mistake, however, to imply that all experiments are cut from the same cloth. There are a number of well-understood threats to internal validity that can compromise results; fortunately, however, most if not all of the problems can be circumvented. *Internal validity* refers to the extent to which a method can establish a cause–effect relationship. The most severe threat to internal validity is a confound. A *confound* exists when more than one thing is unintentionally manipulated in an experiment. The problem is that when two factors are manipulated (e.g., X and Z), any change in the dependent measure (Y) could be caused by a change in X, a change in Z, or some combination of these two factors. Next, I detail nine important confounds that threaten internal validity. Again, I rely on a scenario to illustrate the issues.

Imagine that a schoolteacher is interested in determining whether or not a given reading program will improve reading aptitude in her elementary school classroom. Presume that the program is based on a psychometric theory of intelligence, and the theory makes clear predictions for which students will benefit the most from the program. One possibility would be to administer the program (X) and then measure reading ability (O) as such:

$$\text{X} \quad \text{O}$$

This represents a “one-shot post hoc” experiment, a design recognized as having little (if any) utility in science. The design is severely limited because it does not enable the researcher to claim that X causes (or affects the probability of) O because the design does not eliminate key confounding factors. Sometimes, one-shot post hoc designs are the only possible designs (e.g., the effects of Hurricane Katrina on the Gulf Coast). The key problem here is that the design does not include some kind of control condition, and so it cannot isolate the impact of other factors contributing to the outcome. The important point is that one-shot post hoc designs are problematic because the researcher can never be sure if X or some other factor is the mechanism responsible for producing O.

Being somewhat informed, our teacher opts for a better design. She decides to measure the students’ reading abilities at the beginning of the semester ( $O_1$ ), administer the reading program for 15 weeks (X), and then measure the students again ( $O_2$ ) to determine if they have improved. The overall design is a one-condition pretest/post-test design represented as follows:

$$\text{O}_1 \quad \text{X} \quad \text{O}_2$$

Let’s assume that every student in the class performed better at time 2 than at time 1. The question becomes: can the teacher conclude that the reading program caused better performance? In short, she cannot make a causal claim because there could be a number of confounding factors masquerading as the

treatment effect. Here, I review nine distinct threats to internal validity (for reviews, see [Campbell & Stanley \(1966\)](#) and [Webster \(1994\)](#)).

- *History.* An unlimited number of factors can occur between  $O_1$  and  $O_2$ , in addition to treatment X, that can create a change in the scores. For instance, there may be a national reading campaign that extols the virtues of reading, or the public library might offer incentives for parents to check out books with their children. Any effect of the reading program is perfectly confounded with these factors, and so the impact of that program cannot be separated from them.
- *Maturation.* During the course of the semester, the students will grow and mature at their own rate, both physically and cognitively. Students may improve simply because their brains become more fully developed and their cognitive skills have become sharper with experience. Because this maturation process is confounded with the reading program, again, the impact of the two cannot be distinguished.
- *Selection.* Individuals differ with respect to an array of learned and inherent characteristics. If the individuals in the study are somehow “selected” into groups on a nonrandom basis, then a selection bias exists because personal characteristics may cause a change in reading scores. To illustrate, suppose that a school administrator assigned kids to classrooms based on where they live such that kids from affluent neighborhoods share the same classroom. If so, then socioeconomic status is confounded with the reading program and it may be the true cause.
- *Selection-maturation interaction.* There are sometimes unique effects of maturation processes (related to time) and selection biases (related to personal characteristics) that “interact” in producing an outcome. Suppose that a fair number of “gifted” children are assigned to our hypothetical classroom. The improvement in scores may be caused by the gifted children learning at an accelerated rate (compared to normal children). Thus, the change in scores may have nothing to do with the program per se, but instead it could be caused by the interaction of aptitude and maturation.
- *Testing.* The prior measurement of a dependent variable can sometimes cause a change in the future measurement of that variable. For instance, the very act of measuring reading ability at the beginning of the semester could raise the students’ levels of awareness with respect to their own reading ability. Being more aware, the students may read more often or learn to read more efficiently. Furthermore, if the test is repeated, the students may learn some of the test items. As such, the change in reading ability may be caused by levels of awareness or prior exposure to the test, not the reading program itself.
- *Regression.* Whenever repeated measures are taken, extreme scores tend to become less extreme over time because they move (or regress) toward the group mean. The reason is simple: extreme scores are rare events and

unlikely to occur in succession. In our example, if some children scored very poorly on the first test by random chance alone, they should improve on the second test for the same reason. This change has nothing to do with the program but is a spurious factor fueled by chance fluctuations.<sup>6</sup>

- *Instrumentation.* It is well-known that all measures contain some degree of unreliability that can cause the measurement of the dependent variable to change over the course of an experiment (Thye, 2000). In other words, repeated measures may fluctuate due to random measurement error. When the response item is practiced or well-rehearsed (e.g., responding to “What is your name?”), the random component is small. However, for novel or more difficult responses, the random component is larger. In our example, unreliability in the dependent measure would cause a change in scores unrelated to the true reading ability of the students.
- *Experimental mortality.* During the course of an experiment, some individuals may drop out before the experiment is complete; this is not problematic if those individuals dropped out on a random basis. However, when individuals *selectively* exit an experiment in a nonrandom way, this can create an illusory effect. Imagine that at the beginning of the semester the children vary in terms of reading ability, and over time those with less ability leave because the class it is too difficult or frustrating. At the end of the semester, the average scores on the second test would improve not because the program was effective but, rather, because those children who lowered the original group mean are no longer present.
- *Experimenter bias.* Hundreds of studies show that experimenter expectations can influence subjects’ behaviors in subtle ways (Rosenthal & Rubin, 1978). In a classic study, Rosenthal and Fode (1963) led students to believe they were training either “bright” or “dull” rats to run a maze over the course of a semester. In reality, all students were randomly assigned five ordinary rats with which to work. At the end of the semester, those rats expected to be “bright” objectively learned more than those expected to be “dull.” This occurred because students who expected bright rats actually handled them more frequently and were more patient with them during the training. Thus, student expectations had an impact on rat performance. Another kind of experimenter bias occurs when subjects try to be helpful in the experiment and act in ways to confirm what they believe to be the hypothesis (correctly or otherwise). I address both forms in more detail later.

Despite these threats to internal validity, the logic of experimental design provides a straightforward and powerful solution. Next, I turn to these solutions.

---

6. Strictly speaking, statistical regression to the mean can be caused by numerous factors, including chance fluctuations of subjects’ true score values or random measurement error associated with the instrument.

## VII USING EXPERIMENTAL DESIGN TO RESOLVE PROBLEMS OF INTERNAL VALIDITY

The last mentioned threat to internal validity, experimenter effects, is perhaps the easiest to prevent. In principle, experimenter effects occur because: (1) subjects attempt to confirm or disconfirm the experimental hypothesis; or (2) experimenters unintentionally influence subject behavior. The solution for both problems is to use blinding techniques. A *single-blind experiment* is one in which the subject does not know the true hypothesis under investigation. If the subject does not know the hypothesis, he or she cannot act in ways to confirm or disconfirm that hypothesis. Researchers use a variety of techniques—from simply withholding information to outright deception—to prevent subjects from knowing the hypothesis. Single-blind techniques also prevent other kinds of subject-expectancy biases. For instance, medical research has shown that ingesting an inert substance (e.g., sugar or starch) can make some patients feel better simply because they *believe* they should feel better. Placebo effects are detected by keeping some subjects blind to the experimental treatment.

A more stringent way to prevent experimenter effects is to use a double-blind technique. A *double-blind experiment* is one in which neither the subject nor the experimenter knows the true hypothesis under investigation. Some double-blind studies involve two researchers. One researcher sets up the study knowing the experimental condition; the second interacts with the subjects and/or records their behavior without this knowledge. Other double-blind studies are computer mediated and require only one researcher. Here, the researcher interacts with the subject at the beginning of the study, but then a computer randomly assigns the subject to one of the experimental conditions. In either case, both researcher and subject are blind to the hypothesis. To further prevent experimenter bias, many experimenters use strict protocols that ensure the environment, written and verbal instructions, measures, and so forth are equated for all subjects. Some have said that the best experiment would use perfect clones as subjects (ensuring they are identical) and those clones would never come into contact with the experimenter (ensuring there is no experimenter effect).

Turning to the remaining threats to internal validity, all of these are circumvented through the use of a *true control group*. In our example, the teacher could compare her classroom to another that *did not receive* the learning program. In this case, the second classroom is not a true control group but, rather, a *quasi-control group*. The design is as follows:

<i>Group 1</i>	O <sub>1</sub>	X	O <sub>2</sub>
<i>Group 2</i>	O <sub>1</sub>		O <sub>2</sub>

The inclusion of a quasi-control group helps to eliminate some, but not all, threats to internal validity. For instance, the impact of history and regression to the mean should be equated across the groups because each group experiences

these factors equally and simultaneously. However, because there may be true differences between these groups at the onset of the study, there could be differences in selection and maturation processes across the groups. Thus, although the design is better, there are still factors unrelated to the treatment that could cause changes in the dependent variable.

Without question, the best kind of experiment is the “true experiment,” which involves the random assignment of subjects to conditions, blinding techniques, and the use of a true control group to eliminate threats to internal validity. Within this broad genre, there are many different designs. The simplest is the two-condition “completely randomized design” illustrated next. This is similar to the previous example, except subjects are now randomly assigned to conditions as follows:

<i>Group 1</i>	R	O <sub>1</sub>	X	O <sub>2</sub>
<i>Group 2</i>	R	O <sub>1</sub>		O <sub>2</sub>

Here, the individuals are randomly assigned (R) to one of the two groups before the study begins. Recall that random assignment ensures that subjects are equated on virtually all known and unknowable historic, genetic, physical, personality, and social factors. Thus, at the onset of the experiment, the groups are equated. During the course of the experiment, the treatment group and control group are also equated in the extent to which they are affected by history, maturation, selection, selection–maturation interaction, testing, regression, instrumentation, and mortality. In this context, if the treatment group differs from the control group, it can be reasonably assumed that the difference is caused by exposure to the independent variable. Thus, the combination of random assignment and a carefully tailored control group provides the logical basis for making causal inferences.

More complicated designs involving multiple factors are also possible. Any design with two variables can be called a two-factor design; however, not all are factorial. A true factorial design is one in which each individual subject only appears once. That is, subjects are randomly assigned to one of several experimental conditions defined by the confluence of two or more factors. Imagine that, in our example, students also can receive video instruction (or not) to complement their reading program. To study both factors, the teacher may use a “ $2 \times 2$ ” factorial design in which students are assigned to one of the four combinations. The  $2 \times 2$  design is the most ubiquitous experimental design in the social sciences. The basic structure is shown in [Figure 3.4](#).

Factorial designs offer numerous benefits. First, the design is more informative than the completely randomized design because the experimenter essentially gets two completely randomized designs in a single study. Second, the design is quite flexible. The design shown previously is an independent groups design; that is, both factors are assignable and each individual is randomly assigned to one condition. However, one could also create a factorial

		Reading Program Present	
		No	Yes
Video Instruction Present	No		
	Yes		

FIGURE 3.4 A  $2 \times 2$  factorial design.

design wherein one factor is not assignable (e.g., gender) or both factors are not assignable (e.g., child abuse by gender). Overall, the popularity of the design stems from its sheer strength and flexibility.

Finally, it is worth noting that although the better designs may seem costly in terms of time and money, this should not be a deterrent for those thinking about conducting experimental research. As a practical matter, experimenters often use variations of these basic designs that are more efficient and manageable. For instance, the two-group completely randomized design as presented previously involves four measurements—one for each group before and after the treatment. In practice, experimenters use a variation of this design in which the two measurements before the treatment ( $O_1$ ) are omitted. The logic here is that if individuals are randomly assigned to conditions, the two observations at time 1 will almost assuredly be the same. Omitting  $O_1$  does not alter the basic logic of the design, but it does make the design more elegant and accessible to those who have limited resources.<sup>7</sup>

## VIII VARIETIES OF EXPERIMENTS IN THE SOCIAL SCIENCES

Given the numerous experimental designs, it should come as no surprise that there are different kinds of experiments in the social sciences. For instance, some experiments are designed to test a theoretical prediction, and for these experiments, the measure of success is how well the experimental data conform to that prediction (Martin & Sell, 1979). Other experiments are designed to explore differences across settings and decipher empirical regularities. For these experiments, the metric of success is the ability to detect subtle differences if in fact those differences are real and sustainable. Both brands are informative, but in different ways and for different reasons. Here, I briefly consider these two kinds of experiments with specific emphasis on their purpose and design. This discussion is organized around a central idea: *regardless of the purpose of an experiment, the design of the experiment regulates its validity and overall utility.*

7. In fact, the very first “experiment” I ever conducted as an undergraduate student used this design to test pyramid power on the kitchen counter of my apartment.

## A Theory-Driven Experiments

At one end of the spectrum lies experimental research devised to test scientific theory. The goal here is to compare the outcome of an experiment to a theoretical prediction. For instance, sociologists who attempt to predict social power often use theory to estimate how many points a person will earn during negotiation with another in bargaining games such as those Molm describes in Chapter 9 of this volume. Experiments are then designed that allow people to negotiate, as specified by the theory, and the outcomes of that negotiation are compared to the theoretical prediction. These kinds of experiments are critical in the process of developing and revising scientific theory because they provide a kind of “pure evidence” that either confirms or disconfirms the theory. However, it is important to note that the evidence is only “pure” to the degree that spurious factors cannot infiltrate and contaminate results. The issues of experimental design become critically important in these circumstances.

For theory-driven experiments to be useful, they must be *designed* such that threats to internal validity are thwarted. At face value, the simplest kind of theory-testing experiment occurs when: (1) a theory predicts that a given treatment will produce some effect; (2) that treatment is created in a laboratory situation; and (3) the observed outcome is compared to the prediction. Imagine that a new theory predicts that when males and females disagree, the males change their opinion exactly 42% of the time and stay with their opinion 58% of the time. Under this arrangement, we have a theoretical prediction (P), a treatment (X=mixed-gender disagreement), and an outcome (O=42% opinion change). The design of the experiment is as follows:

P            X            O

Clearly, the reader should recognize that this design is simply a variant of the “one-shot post hoc” model outlined previously and, as such, is vulnerable to all of that design’s problems and pitfalls.

Two issues render this kind of experiment problematic. First, all of the spurious factors that threaten the internal validity associated with the one-shot post hoc design (history, maturation, etc.) also jeopardize the integrity of this design. Second, ironically, although this experiment is designed to “test theory,” the outcome is equivocal with respect to theory. Assume the experiment is conducted and the results correspond perfectly to the prediction that males change their opinion exactly 42% of the time. Does this mean the theory is confirmed? It is difficult to say. Although the data confirm the prediction, the reason for this may have nothing to do with the processes asserted by the theory. It could be that males *always* change their opinion 42% of the time, regardless of the status of their partner. In other words, the outcome may correspond to the prediction for reasons that have virtually nothing to do with the theory. The theory claims that X produces the outcome, but because the preceding design does not allow comparisons to situations without X, the confirmation status of the theory is obscure.

To conduct theory-driven research in its most instructive form necessitates certain kinds of controls. There are at least two varieties of control normally employed, alone or in conjunction, to sharpen the meaning of experimental data. The first and most straightforward type of control is the *control group*. Referring back to the preceding example, the researcher can randomly assign some males to interact with females and some males to interact with males who are status equals. Assume the researcher finds that males interacting with females change their opinion 42% of the time as predicted, whereas males interacting with status equals change their opinion 53% of the time.<sup>8</sup> Such a finding speaks more to the veracity of the theory because it suggests that opinion change (O) is contingent on gender (X), as the theory predicts.

Second, researchers sometimes compare findings across studies as a kind of pseudo-control group in the absence of a true control group. For instance, countless studies in economics and sociology find that when two people negotiate in a dyad, neither has an advantage and they tend to profit equally. Furthermore, this finding is predicted by numerous sociological (Emerson, 1981; Willer, 1999) and economic (Nash, 1950) theories. Given the breadth of these theories and their evidentiary basis, such findings are now taken for granted. Researchers can use these prior experimental sessions as a kind of “control” or “baseline” with which to compare new research findings. Often, this is done as a practical matter when a finding has been replicated to the point that yet another demonstration provides little or no new information.

## B Empirically Driven Experiments

At the other end of the spectrum are experimental studies targeting empirical problems. For instance, the most common use of experimentation in medicine, biology, agriculture, and education is to detect differences across two or more treatments. The experimental food scientist may attempt to understand how different levels of refrigeration accelerate bacterial growth on Atlantic shrimp; the experimental educator may want to understand whether computer-mediated instruction improves student retention. In both cases, the purpose of the experiment is to identify possible differences across or between treatments. These experiments require the same kinds of control over spurious factors as do theory-driven experiments for the same reasons discussed previously, so those points are not duplicated here. However, because the purpose of the experiment is to make inferences to a larger empirical population, the empirically oriented experimenter must confront another set of problems.

In empirically driven experiments, the researcher is almost never interested only in the properties or behavior of the specific experimental units under investigation—that is, these shrimp or these students. Rather, the goal of

---

8. For this example, I presume the 11% difference between 42 and 53% has statistical and practical significance.

the experiment is to make inferences about the population from which these units originate—something akin to all shrimp stored at this temperature or all students exposed to this style of teaching. In other words, the experiment is a device used to make population inferences from sample data. In this business, there are two kinds of mistakes that can be made, and these have different consequences for the conclusions that can be drawn.

The first mistake occurs when the researcher claims a false positive. That is, when the study finds a difference in the sample data but there is no true difference in the population. Statisticians refer to this as *Type I* statistical error. Type I error is regulated by the experimenter because the experimenter sets the alpha level (or Type I error rate) for the statistical test. In a statistical sense, this controls how frequently one rejects the null hypothesis (i.e., how often one finds differences across treatments) when that hypothesis should fail to be rejected (i.e., there are no differences across treatments). There are certain principles that guide how scientists set the Type I error rate. For instance, when an experiment bears on the benefits of a treatment, most empirically oriented experimenters adopt a more stringent Type I error rate, such as 5 or 1%. The idea here is to be skeptical about the claimed benefit of a treatment and demand fairly rigorous proof of it. On the other hand, when the experiment bears on the negative or harmful impact of a treatment, such as the detection of cancer, most researchers use a more liberal Type I error rate, such as the 10% level. The corresponding logic is to be more sensitive to potentially harmful effects.

The second kind of mistake is to declare a negative result when that conclusion is unwarranted. A *Type II* error occurs when one finds no difference between treatment groups but true differences exist in the population. The inverse of Type II error, *statistical power*, is the probability of detecting a difference in your sample when that difference is real in the population. Power is affected by a number of factors, including the size of the experimental groups, the Type I error rate, the variability of the phenomenon, and the magnitude of the differences between treatments. In many ways, the lack of statistical power (Type II error rate) is the most serious problem confronting all experimenters. This is because, in principle, virtually any experimental treatment will have *some effect* on the dependent variable, however small. That is, following a manipulation, the probability that a treatment and control group are *exactly* the same on the outcome measure is 0. The question is: does the experiment have adequate power for the researcher to detect that effect correctly? When designed with adequate power, empirically driven experiments are refined tools for understanding the empirical world.<sup>9</sup>

---

9. Type I and Type II errors can also be made in theory-driven research, but with potentially more severe consequences. Importantly, these errors will ensure that true theoretical assertions die and false ones survive.

## IX EXTERNAL VALIDITY AND ARTIFICIALITY

Because only a relatively small number of social scientists employ experimental methods, it may be worthwhile to address some general misconceptions regarding the role of experimentation in the process of producing general knowledge. Perhaps the most misunderstood issue surrounding experimentation is that of external validity. *External validity* refers to the degree that experimental findings hold for other persons, in other locations, at other times. Recall from the introduction that [Babbie \(1989\)](#) and other writers of methods texts ([Silverman, 1977](#)) frequently claim that because laboratory studies are conducted in “artificial” environments, their findings do not generalize to other circumstances. Overall, the two kinds of experiments detailed previously (theory driven vs. empirically driven) have different purposes and deal with the issue of external validity in different ways. To foreshadow, the purpose of theory-driven experiments is to provide one kind of test for a scientific theory. For these experiments, *theory*, not empirical *generalization*, connects the experiment to naturally occurring phenomena (see also [Lucas, 2003](#); [Mook, 1983](#); [Thye, 2000](#); [Webster & Kervin, 1971](#)). For empirically driven experiments, *statistical inference* provides the link between the experimental outcome and the larger population. Next, I address each kind of generalization.

### A Theory-Driven Experiments

The most frequent argument against the use of theory-driven experiments is that they have low external validity. A related argument is that the laboratory is an artificial environment and so it has limited utility to understand events in the real world, as illustrated by the quotes that open this chapter. It is true that laboratory results cannot *directly generalize* beyond the laboratory. However, they can empirically document theoretical principles and rule out alternative explanations in ways that are useful for understanding aspects of “real-world” events with appropriate qualification (see [Zelditch, 1969](#)). It is the case that some who conduct experiments in the social and behavioral sciences engage in rhetorical practices that suggest their results are *directly generalizable* without theory. This only weakens the apparent value of experimental research in the eyes of those who employ other research methods.

Nevertheless, discounting or altogether rejecting laboratory methods because they do not directly generalize indicates a misunderstanding of the role that these kinds of experiments play in the research process. Issues of theory testing and external validity are no different for sociology than for any other field employing laboratory methods, and yet: (1) all laboratories in all sciences create artificial conditions; (2) all sciences try to explain real-world phenomena; but (3) experimental methods are embraced proudly, unquestioningly, and highly successfully by other sciences.

Used properly, theory-driven experiments do not produce phenomena whose descriptions may be generalized to natural settings. *These kinds of experiments*

*test theories.* Unfortunately, whenever experimentalists in psychology, social psychology, and sociology operate with less than explicit, *a priori* theories—which sometimes they do—their experiments disengage from their proper theory-testing function. When not actually testing hypotheses derived from theories, it may be tempting to infer that experiments are intended to serve as microcosms of real-world contexts, permitting the generalization of experimental findings to those natural settings. Even while some experimentalists encourage such interpretations, there is no basis for doing so (see Thye, 2000).

As Webster and Kervin (1971) explain:

*The proper use of the laboratory permits no direct generalization of laboratory results to the outside world; the only permissible connection between the two is the theory. Therefore the artificiality of the laboratory setting is an irrelevant issue when one is speaking about results in the natural environment. What is relevant is whether the natural environment ever contains instances which approximate situations described in the scope conditions of the theory. (p. 269)*

When a laboratory experiment is used to confirm or disconfirm a theory, findings in that setting are as relevant to that theory's truth or falsity as findings gathered from any other setting. Moreover, given the degree of control over extraneous factors that the laboratory affords, the relevance of a given experimental test to a specified theory usually is even greater than a parallel test in a natural setting. Indeed, Webster and Kervin (1971) argue that what may be needed in sociology is *more* artificiality in the form of experimental research because the history of science has demonstrated it to be the most efficient way to promote theoretical development. The advantage of the experiment lies, to a great extent, in its ability to filter the noise that may interfere with drawing sound inferences.

In summary, the purpose of theory-driven experimentation is to *create* a testing ground that falls within the scope of the theory under consideration. Perhaps Mook (1983, p. 380) said it best when he stated, “We are not *making* generalizations, but *testing* them.” Experimental conditions are only relevant as they relate to the highly stylized abstract theories they inform. In turn, those theories may be useful for understanding a range of phenomena. In physics, for example, experimentally informed theories make claims under conditions known never to occur in nature, such as “in a perfect vacuum.” However, these theories and their experimental foundation provide key insights for resolving real-world problems. Using theory to understand various problems across diverse domains is a process of *theoretical inference*.

## B Empirically Driven Experiments

The issue of generalization is somewhat different for experiments designed to detect empirical differences. Such experiments are not guided by theory; instead, they tackle practical problems and issues (e.g., which drug reduces

pain, which therapy increases flexibility, and which grain is most resistant to drought). Here, it is not theoretical inference but *statistical inference* that allows one to forge connections between sample data and the larger population from which that sample came. Fisher (1935) was the first to formalize statistical models that allowed such inference.<sup>10</sup> These models enable one to claim, with some probability, that differences between experimental treatments are real and exist in the population. The issue of statistical generalization is no different for experimental work than for any other quantitative area. In all cases, correct inference is governed by probability theory and the statistical power associated with the hypothesis test.

Although critics assert that experimental findings do not directly generalize to other settings, only a few researchers have actually taken the time to check. For instance, Dipboye and Flanagan (1979) examined the content of empirical articles from major psychological and organizational journals over several years to determine if field research is broader than the typical laboratory study. Contrary to popular belief that field research is more representative, they found studies in the field to be just as narrow as those in the laboratory in terms of the subjects, behaviors, and situations under investigation. More directly, Locke (1986) examined research findings from industrial organizational psychology, organizational behavior, and human resource management. These are fertile testing grounds because the prevailing theories in these areas: (1) are heavily informed by laboratory research; and (2) ultimately guide working conditions in businesses and organizational environments. The question is whether or not the laboratory findings reproduce themselves in the business world. It turns out they do. Time and again, Locke and colleagues found remarkable consistency between the field and the laboratory. Locke writes, “Both college students and employees appear to respond similarly to goals, feedback, incentives, participation, and so forth, perhaps because the similarities among these subjects (such as in values) are more crucial than their differences” (p. 6). Thus, despite the prevalence of the claim, the data suggest that many laboratory findings generalize to field settings.<sup>11</sup>

## X CONCLUSION

I began this chapter with the goal of outlining the philosophical and logical foundations of experimental methods in the social sciences. For those of us engaged in the business of designing and executing experiments, the advantages this method affords in terms of promulgating scientific theory and aiding

10. In fact, the statistical “*F* test” that accompanies modern ANOVA models is named in honor of “Fisher.”

11. It may not be obvious, but the structure of generalization in empirically driven experiments is the same as in theory-driven experiments. In both cases, research findings from one setting must be thought about more *abstractly* before those findings can be mapped onto and empirically checked in another setting.

empirical exploration are unmistakable. However, we are still a small (although growing) segment of practicing social scientists. Critics are decidedly more abundant, and vocal, as they galvanize around issues of random assignment, artificiality, external validity, and experimental utility. In some ways, it is difficult to conceive that the experimental strategy—the touchstone of scientific inquiry in physics, medicine, biology, chemistry, and so on—was argued to be in critical condition approximately a decade ago. In truth, experimental methods in the social sciences serve the same function as they do in other scientific arenas. *Experimentation is the best known way to test the veracity of scientific theories, eliminate alternative explanations, engineer novel solutions to practical problems, and provide clues to causal inference.*

In closing, the experimental method is the *sine qua non* of scientific inquiry, spanning disciplines from particle physics to aerospace exploration and everything in between. The ubiquity of the experimental method likely stems from its multifaceted nature and utility. The method provides a powerful yet elegant way to: (1) build, revise, and sometimes dismantle scientific theory; and (2) explore the empirical world and seek practical solutions to problems not well understood. The two processes are intertwined: theories are used to understand practical problems, and such problems are the inspiration for theory development. The interactions of these two domains fuel and propel the scientific process. As long as both are present—theories to test and real-world problems to solve—the experimental method will be alive and well.

## ACKNOWLEDGMENTS

I thank the National Science Foundation (SES-0216804) and the University of South Carolina for supporting my research. I thank the editors, Murray Webster and Jane Sell, for their insightful comments and suggestions. I am also grateful to Lisa Dilks, Kyle Irwin, Edward J. Lawler, Tucker McGrimmon, Barry Markovsky, Lala Steelman, and Jennifer Triplett, who helped me identify a humbling number of errors in the manuscript. Regretfully, none of us can be sure that we got them all.

## REFERENCES

- Aristotle. ([340 Bc] 1947). *Metaphysics*. Cambridge, MA: Harvard University Press.
- Babbie, E. (1989). *The practice of social research*. Belmont, CA: Wadsworth.
- Bunge, M. (1979). *Causality and modern science*. New York: Dover.
- Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.
- Cohen, B. P. (1989). *Developing sociological knowledge: Theory and method*. Chicago: Nelson-Hall.
- Cohen, B. P. (1997). Beyond experimental inference: A decent burial for J. S. Mill and R. A. Fisher. In J. Szmata, J. Skvoretz, & J. Berger (Eds.), *Status network and structure: Theory development in group processes* (pp. 71–86). Stanford, CA: Stanford University Press.

- Cohen, M. R., & Nagel, E. (1934). *An introduction to logic and scientific method*. New York: Harcourt Brace and World.
- Davis, J. A. (1985). *The logic of causal order*. Newbury Park, CA: Sage.
- Dipboye, R. L., & Flanagan, M. F. (1979). Research settings in industrial and organizational psychology: Are findings in the field more generalizable than in the laboratory? *American Psychologist*, 34, 141–151.
- Emerson, R. (1981). Social exchange theory. In M. Rosenberg & R. H. Turner (Eds.), *Social psychology: Sociological perspectives* (pp. 30–65). London: Transaction.
- Fisher, R. A. (1935). *The design of experiments*. London: Oliver & Boyd.
- Fisher, R. A. (1956). *Statistical methods and scientific inference*. Edinburgh, UK: Oliver & Boyd.
- Galileo, G. (1954). *Dialogues concerning two new sciences*. New York: Dover (Original work published 1636).
- Hage, J., & Meeker, B. F. (1988). *Social causality*. Boston: Unwin Hynman.
- Harré, R., & Secord, P. F. (1972). *The explanation of social behavior*. Oxford: Oxford University Press.
- Hempel, C. (1965). *Aspects of scientific explanation*. New York: Free Press.
- Henshel, R. (1980). Seeking inoperative laws: Toward the deliberate use of unnatural experimentation. In L. Freese (Ed.), *Theoretical methods in sociology: Seven essays*. Pittsburgh, PA: University of Pittsburgh Press.
- Hume, D. (1955). *An inquiry concerning human understanding*. New York: Bobbs-Merrill (Original work published 1748).
- Lieberson, S. (1985). *Making it count: The improvement of social theory and research*. Berkeley, CA: University of California Press.
- Lieberson, S. (1991). Small N's and big conclusions: An examination of the reasoning in comparative studies based on a small number of cases. *Social Forces*, 70, 307–320.
- Locke, E. A. (1986). Generalizing from laboratory to field: Ecological validity or abstraction of essential elements. In E. A. Locke (Ed.), *Generalizing from laboratory to field settings* (pp. 3–9). Lexington, MA: Health.
- Lucas, J. (2003). Theory-testing, generalization, and the problem of external validity. *Sociological Theory*, 21, 236–253.
- Martin, M., & Sell, J. (1979). The role of the experiment in the social sciences. *Sociological Quarterly*, 20, 581–590.
- Mill, J. S. (1973). *A system of logic, ratiocinative and inductive*. Toronto: University of Toronto Press (Original work published 1872).
- Mook, D. (1983). In defense of external invalidity. *American Psychologist*, 38, 379–387.
- Nash, J. (1950). Equilibrium points in *n*-person games. *Proceedings of the National Academy of Science, USA*, 36, 48–49.
- Paulos, J. (1988). *Innumeracy: Mathematical illiteracy and its consequences*. New York: Vintage.
- Rosenthal, R., & Fode, K. L. (1963). The effect of experimenter bias on the performance of the albino rat. *Behavioral Science*, 8, 183–189.
- Rosenthal, R., & Rubin, D. B. (1978). Interpersonal expectancy effects: The first 345 studies. *Behavioral and Brain Sciences*, 3, 377–386.
- Russell, B. (1913). On the notion of cause. *Proceedings of the Aristotelian Society*, 13, 1–26.
- Silverman, I. (1977). *The human subject in the psychological laboratory*. New York: Pergamon.
- Thye, S. (2000). Reliability in experimental sociology. *Social Forces*, 74, 1277–1309.
- Waldrop, M. M. (1992). *Complexity: The emerging science at the edge of order and chaos*. New York: Simon & Schuster.
- Webster, M. (1994). Experimental methods. In M. Foschi & E. Lawler (Eds.), *Group processes: Sociological analyses* (pp. 43–69). Chicago: Nelson-Hall.

- Webster, M., & Kervin, J. (1971). Artificiality in experimental sociology. *The Canadian Review of Sociology and Anthropology*, 8, 263–272.
- Willer, D. (1999). *Network exchange theory*. Westport, CT: Praeger.
- Willer, D., & Walker, H. A. (2007). *Building experiments: Testing social theory*. Stanford, CA: Stanford University Press.
- Zelditch, M., Jr. (1969). Can you really study an army in the laboratory? In A. Etzioni (Ed.), *A sociological reader on complex organizations* (pp. 531–539). New York: Holt, Rinehart & Winston.

## Chapter 4

# Training Interviewers and Experimenters

Robert K. Shelly  
*Ohio University, Athens, Ohio*

### I INTRODUCTION

Preparing research assistants to carry out an experiment that will lead to an informative test of research hypotheses involves several important elements. These require substantial thought on the part of the investigator to design the study, time to develop materials and train experimenters, and the recognition that supervision is important. Working with the students we employ for these roles often requires that we take on the role of a socializing agent similar to that of a parent when we first involve them in our work. Recognizing and executing this complex pattern of identity management sometimes requires great patience and the willingness to exercise restraint in the face of monumental errors.

As I develop the plans for a program to develop skills and abilities in new experimenters and interviewers in the following pages, it is important to remember that this situation involves a relationship in which both parties are learning how the work of the experiment will be carried out. I have worked with a large number of students as research assistants throughout my career. They have each had different levels of ability and prior preparation when we started our work. Many have had a course in research methods and nothing more; others have had extensive experience working with other investigators. Some have developed new skill sets that have led them to be able to creatively contribute to my research program. Others have been effective research assistants but for various reasons decided that their contributions will be limited to carrying out the tasks I set for them. As an investigator, I value both types of experiences, even though I would like to think that every student I train will become a “junior me” and “take over the business” of my research program when I stop working on it.

### II PRELIMINARIES

The first step in any training program is to determine what the research assistants are to do—in other words, the job description. This is often set by the type

of data one wants to collect. Is the study designed to collect open interaction data with video recordings of groups discussing a problem for an hour? Will the data then be transcribed or will the data be coded directly from the video record in the 21st century equivalent of Bales' (1950) real-time coding sessions? Good presence in presentation and the ability to "perform" to a script may be sufficient for collecting data in this situation. If assistants are to code these types of data, then a more thorough training program that addresses how to recognize behavioral categories, their start and stop events, and the passing of turns to another group member may be appropriate. Part of the training program should include details on how to recognize and code nonverbal and paraverbal behavior if this is part of your research interest.

Skills needed to conduct an open interaction study are different from those needed to conduct a study employing computer-mediated (pseudo) interaction in which the data are stored directly to a file and minimal coding is needed to pursue an analysis. In the former, the research assistant may need to start and stop the session and make sure the data are properly stored. In the latter, the research assistant may need computer skills that will allow him or her to troubleshoot problems that develop in the course of the study. The possibility that errors may be introduced by equipment failures and the ability to quickly redress these faux pas may save experimental sessions and reduce the need to schedule additional data collection.

We often collect questionnaire and interview data in the course of an experiment. This is done frequently to determine if participants meet the scope conditions of the theory. The research assistant needs to understand the purposes behind this effort and appreciate that it is an integral part of the experiment. Skills required here are highly variable. In some cases, it may be sufficient if the assistant is able to collect, sort, and store paper-and-pencil instrumentation. In others, it may be desirable for the assistant to conduct a complex open-ended interview that is recorded and then review the recording and paper-and-pencil records to make a decision about whether or not to include the subject in the pool of "good" data. Decisions about these issues affect the training program one implements.

Recruiting assistants depends on available pools of personnel. I frequently find that courses I teach serve as a fertile ground for interesting students in the chance to participate in data collection. This is particularly true for courses in research methods and in my substantive area of interest in group processes. Both environments have proved to be fertile grounds for cultivating and recruiting future research assistants. Because I teach these courses at the undergraduate and graduate levels, I recruit potential research assistants with varied levels of background from my classes. This can present training problems if I am working with students who have diverse backgrounds when I start an experiment.

The other option is to take potluck with assigned graduate assistants from the department pool. This has proved to be less satisfactory because students are often assigned based on availability and not on interest. Assistants assigned

in this way may require more motivation and supervision. Occasionally, I have had to alter the nature of the assignment if the student proves to be unsatisfactory as a research assistant. In general, working with the department chair or director of graduate studies so they understand your interests and needs may be very beneficial. Requests for particular students or sets of skills are more likely to be honored if the responsible person understands the requirements of your projects.

It is also important to train assistants in research ethics. An understanding of federal and institutional policy regarding the treatment of human subjects is critical for the entire research staff. A variety of means are available to accomplish this end. My own institution has an online training program that qualifies an individual for certification as having completed the appropriate training. I usually try to provide some preparation for research assistants in this area, but it is highly variable and depends on the roles I expect them to play. If the person is only doing data entry, I emphasize the importance of keeping data confidential. If the person is carrying out experiments that involve deception, I emphasize the importance of careful and thorough preparation for the study, the reasons for the deception, and the importance of careful debriefing. For research assistants whom I expect to be heavily involved in my projects, such as those that involve longer term relationships or the expectation that the assistant will be involved in developing the protocols for the study, I ask that they become familiar with institutional policy, federal policy, and the appropriate professional code of ethics.

Another area of interest that affects the training program is the nature of the raw material with which one starts. In many ways, this is critical. Beyond the obvious point that we want someone who is reliable, conscientious, and at least a bit compulsive, good preparation is essential. I prefer to work with students who have completed a general research methods course that covers the major techniques of data collection as well as some discussion of scope and initial conditions and also data analysis. This preparation has the effect of creating a common language of discourse so I do not have to teach them how to think about the role they play. To borrow an analogy from the theater, I do not expect Paul Newman as my lead, but I hope the aspiring experimenter has at least been to acting school.

### **III EXPERIMENTS AS THEATER**

I wish to continue the analogy with the theater for a time by pointing out that the experiment is a dramatic event. As investigators, we often plan our studies so as to restrict the interaction to a relatively narrow domain of activity. We are interested in the verbal and nonverbal activity of our participants as they try to solve problems in open interaction or make decisions about offers and counter-offers in a negotiation. Our objective in creating experimental situations is very much like the effort in the theater to create a suspension of disbelief. We wish to convince our participants that the situation they confront is real, that the other

participants are genuine in their actions, and that consequences of actions are real. All of these elements of a good experiment require careful planning and a good cast to carry them out.<sup>1</sup>

The mode of data collection is particularly important in designing the training program. If the project is a vignette study that requires collection of questionnaire data in classrooms or laboratories with minimal interaction, it may be sufficient to work with your researcher to read a prepared script. If the study requires that the research assistants serve as an experimenter or as a confederate who plays the role of a “real subject,” you may need to train them to enact these roles. This training should emphasize that assistants must follow the script and that the experience must be real for the subject. The experiment is designed to create a believable experience for the subject. The research staff must accomplish this goal. Simply reading the script is not sufficient; managing emotional tone and impressions given off as material is presented to subjects are critical elements in creating a successful data collection effort.

The theatrical analogy includes the creation of space for the experiment, the selection and proper placement of props to carry off the cover story, a script that details actions and identifies what to do if unexpected contingencies develop, knowledge of roles that must be filled, and a cast that has been well rehearsed. Space for the experiment can be almost any environment that meets the needs of the study. I have run discussion group studies in a physical anthropology laboratory with a cased skeleton in the wings,<sup>2</sup> in an archeology laboratory with tools and artifacts lying about, and in classrooms with all of the leftovers at the end of the day scattered around the room. Other investigators I have shared these stories with tell of similarly Spartan or unusual accommodations for their work. The key is to recognize how to use the space, arrange props, and create a believable social experience for the participant(s).

It is also advisable to create a sense of “onstage” and “offstage” for the research staff by designating areas where assistants can be away from the experiment and yet remain available for their parts in the study. It is important that distinctions about where “backstage” and “frontstage” behavior are appropriate be well understood and enforced. Inappropriate activity in either onstage or off-stage areas can have negative consequences for the experimental session where the event occurs or jeopardize the entire study if the breach is egregious.

The risks of inappropriate “frontstage” behavior can often be managed with rehearsals and practice sessions, but “backstage” problems may not appear until sessions are actually carried out with live subjects. For instance, inadequate soundproofing in a laboratory space may require that the staff remain silent while the subjects engage in a discussion or work on a computer for the experimental sessions. It is also advisable to train your staff in appropriate language to

---

1. Provocative discussions of issues relating to casts and production in theater are contained in [Willis \(1994\)](#).

2. The skeleton was behind a moveable partition, out of the view of the participants in the study.

be employed when discussing the study with subjects and one another in public. For instance, references to subjects should be as “participants.” Emphasizing the importance of a professional demeanor is also important as experimenters learn how to conduct sessions.

Props include a variety of objects. They may include writing instruments, voice recorders, video cameras, supplies such as audio or video media, research questionnaires for recording responses in interviews, and the mundane such as tables and chairs. Management of these items and training in their use should be a part of any preparation to carry out an experiment. The training program should include discussions of what to do if equipment fails and the possibility that data may be lost due to failure to properly record information about sessions, conditions, and days and dates a session was completed. Finally, some words about costume. Most experiments involve the presentation of information from a legitimate researcher. Because legitimacy is often tied to position and appearance, it is likely that the experimenters will have to dress as “professionals.” I have found that some simple direction about presenting a neat or professional appearance is often sufficient to encourage experimenters to leave the ragged T-shirt and cutoff jeans at home and appear in clothes that would get them seated in most family restaurants in the United States.

Developing a script for the experiment, and its rehearsal, is a significant part of any preparation for a study. The exact content of the script will depend on the experiment and data being collected. Rehearsals can take a variety of different forms. One could video practice sessions of the study, review audio recordings of the exit interview and debriefing, or observe rehearsals in real-time. Whichever of these techniques one chooses to employ, feedback early and often is important. Theatrical directors employ notes to their actors to encourage character development and adherence to the script. You may wish to develop a similar strategy about rehearsals. Formal notes have the advantage of requiring you to identify specific behaviors that must be corrected, and they may form the basis of action if you need to terminate a relationship with an experimenter.

The degree of prior experience and preparation of your experimenters will affect the need to rehearse and refine the performance for participants. This may actually be a two-edged sword, however. Experienced assistants may believe they have little to learn but may have bad habits to unlearn from your point of view. Training and rehearsals are the places to solve these problems; losing data because of bad behavior by an experimenter is not acceptable.

## **IV PREPARING THE ASSISTANT FOR THE ROLE**

Determining how to train an assistant depends on the nature of the data collection task. The development of skills needed for supervising an entire data collection enterprise is more time-consuming than training for only one or two tasks. In addition, how much to reveal to a new experimenter about the goals of an experiment is always a question with technical and substantive import.

I discuss these issues in turn, beginning with theoretically important issues on which beginning researchers need to be prepared.

The experimenter should understand the role of scope conditions in the research so that he or she can effectively make decisions about the quality of data from a particular experimental session. For instance, expectation states experiments require that participants are collectively focused on arriving at the best solution to the group task. If they are not, the data they produce is suspect and generally not included in the set for analysis. Learning how to determine whether an individual is collectively focused is a key part of the preparation of the experimenter. Are decisions to be made on the basis of volunteered information or at the end of a detailed interview conducted at the conclusion of the experimental session? What decision criteria are to be employed to include or exclude a participant in the final data set?

All these points require that the experimental assistants have a grasp of the theoretical points under investigation. This frequently requires that they become familiar with research literature that many graduate and almost all undergraduate students will find unfamiliar and daunting. Careful selection of materials and a planned reading program are frequently valuable in preparing a potential assistant in this aspect of conducting studies. The extent of this preparation will depend on how you expect the person(s) to work with you on your research. If the students are to become a part of an ongoing research team, extensive preparation is in order. If they are to serve only as experimenters for a single study, then brief, targeted materials may be sufficient.

The creation of a reading program for training assistants will depend on the role you wish to have them fulfill. Students who are to collect and enter questionnaire data in a vignette study may be exposed to very little research literature if that is their only role. If the student is to carry out coding of open interaction data, it may be desirable to expose him or her to some material on content analysis and coding of interaction. The particular work chosen will depend on your objectives. If students are expected to carry out experimental sessions, code and enter data, and make decisions about including and excluding subjects from the final data analysis, more extensive reading lists may be in order. Decisions about how far to carry this education will depend on the possibility that experimenter effects are a concern.

Any experimental study has a set of scope conditions and a set of initial conditions that define the social situation within which the subject acts. Scope conditions define the situation of action such as collectively orientated group activity aimed at solving a task. Initial conditions specify the relationships between social actors in the situation such as initial advantages and disadvantages group members have with respect to resources or status in the group. Student experimenters should have a grasp of initial conditions so they can determine if subjects believed they were sufficiently similar to or different from one another to satisfy the conditions of the experimental design.

In expectation states experiments, participants are informed that they occupy the same or different status as their partner in the study. Learning how to

assess whether this information was internalized by the participant is a key element in determining whether or not to include data from a particular participant for analysis. Did the participant understand the information that was presented? Did he or she accept and believe it? Did he or she act on it? These are all questions the experimenter must answer to reach a decision about including data for a particular individual in the final data set. Explicit criteria for assessing a post-session interview help, but prior training will produce higher quality decisions.

My reading list for novice experimenter trainees includes Bernard Cohen's (1989) *Developing Sociological Knowledge*. The discussion of initial and scope conditions, the use of indicators as measures for concepts, and the linkage between these key elements of testing hypotheses is one of the best available. If students have difficulty understanding or appreciating the material, I often consider them for more limited roles in the research enterprise.

Providing a novice experimenter with explicit and detailed knowledge about hypotheses under investigation may be more problematic. Ideally, a well-prepared research assistant will have a detailed understanding of the measures, hypotheses, and expected values of dependent variables in the study he or she is conducting. Robert Rosenthal (1976; Rosnow & Rosenthal, 1997) and colleagues raise questions about introducing artifactual biases into the data collection process when an experimenter has this detailed knowledge. Such biases may contaminate data and lead to false acceptance or rejection of research hypotheses.

An impressive array of studies during approximately the past 40 years has detailed how human experimenters produce adverse effects on outcomes of their studies with subtle, often difficult to detect communication to research participants. This work is variously accessible because many of the books are out of print, but you must be aware of the problems inadvertent, nonverbal, and paraverbal communication may create as participants are instructed, supervised, and interviewed by an experimenter. This problem is by no means limited to beginners working on their first experiment. Investigators will be aware of these problems and prepared to train personnel so they may be minimized or avoided.<sup>3</sup>

The possibility that experimenters may have unintended effects on results is a significant issue for social science experiments. Recognizing this should not paralyze us, however. The use of multiple experimenters and the replication of conditions across experiments will reduce the risks of making incorrect decisions about the quality of data from experiments. One solution often advocated to the possibility that experimenters may inadvertently affect outcomes is the use of double-blind experiments. These may be appropriate in some circumstances but not others. When the experiment requires an interview or questionnaire specific to particular initial conditions, keeping the information from the

---

3. *People Studying People: Artifacts and Ethics in Behavioral Research* (Rosnow & Rosenthal, 1997) is a remarkably readable treatment of the issues in this area and should be on the reading list for new experimenters as they begin their training.

experimenter is not possible. In such circumstances, it may be appropriate to ensure variation in assignment of experimental sessions to experimenters so that an assessment of possible experimenter effects can be carried out.

As a research assistant develops skills and takes on responsibilities during the conduct of investigations, you must confront the issue of the extent to which the person is contributing intellectually to the project. As mentioned previously, I try to prepare research assistants as if they will take up my research program. This often involves training not only for the limited roles of conducting experiments but also for the intellectual tasks of contributing to the development of the research program. Ideally, such preparation takes on the qualities of an apprenticeship program, with roles and responsibilities increasing as the assistant gains experience. The ultimate goal is for the student to be able to develop his or her own research project that answers important theoretical questions based in your program. Often, this requires at least a year of work with you on a project before the student is ready to proceed on his or her own. Guiding the student through this process requires a well-thought-out plan on your part. Students do not develop the skills to accomplish this end until they have “practiced” on several projects. The goal should be to identify and work with the various elements that go into translating a theoretical idea into a study, identifying how to develop the study, and implement the logistics of the study.

## V TRAINING FOR SPECIFIC TASKS

### A Conducting Experimental Sessions

When working with individuals who have no experience in collecting data or who have not previously conducted experiments, I design a ladder of opportunities to prepare them to supervise the experimental sessions. The first set of experiences involves simple tasks such as telephone scheduling of participants as the first step. This generally allows them to learn to follow a script and interact with participants; by asking for feedback about their experience, it offers an opportunity for me to assess their skills. Depending on the nature of the experimental protocol for the session, I then move the assistant into a more responsible role by having him or her participate in sessions I conduct. Usually, this involves asking the assistant to conduct postsession interviews, serve as a secondary experimenter if the role exists for the study, or act as a greeter if multiple locations are in use for participant arrival. This process usually does not require many sessions to provide some rudimentary training so the new assistant can move on to more complex tasks.

In addition, I often ask the trainee to take part in the study as if he or she were a naive participant. Not only does this increase the understanding of the trainee but also it leads to important information about the effectiveness of the manipulations and the clarity of instructions. Questions and comments of the new experimenters are often more informative for the experimental protocol than comments gleaned from postsession interviews of naive participants. This information can

highlight problems of implementation because the trainee has developed some knowledge about the intent of the study and is often aware of the expected effects of the experimental manipulations. The trainee is also less likely to be intimidated by the experimental setting than the typical naive participant. The earlier in the sequence of collecting data this can be accomplished, the better. Information gleaned from such an experience at the pretest stage of a study is much more valuable than information secured as data collection is concluding.

Once the assistant has completed all of these steps, I feel comfortable working him or her into the overall rotation to run sessions. I sometimes continue to run sessions myself and compare notes for a few days to ensure the data collection is going as intended. It is also valuable to plan some retraining from time to time to keep research assistants sharp and to ensure they are following the protocol as expected if the data collection run is to extend for a long period of time, such as across several academic terms. Some occasional monitoring of sessions may be in order as well. This should be done with the knowledge of the assistant to ensure he or she understands that it is done to ensure a quality execution of the protocol.

## B Coding Interaction Data

Learning to code open interaction data is in many ways much easier than it was when I began working in this area as a graduate student. We were still using Bales Interaction Recorders (1950) and began training by tapping pencils on the table to signify acts as we observed groups through a one-way mirror. This had the advantage of creating some reliability in our identification of events, but we still had to classify the activity. Interaction coders today have the advantage of being able to replay a video record to determine if the behavior they are observing is the initiation of a new act, if it is in response to an action of a specific other member of the group, and how it should be classified according to the coding scheme in use in your investigation. One key issue to confront is whether you wish to code the data from a video record or from a transcript prepared from the record. Both have advantages and disadvantages and also different training demands.

Coding directly from the video record has the advantage that you can employ codes for verbal, nonverbal, and paraverbal behavior. Techniques for implementation of this scheme are relatively easy to train with software that is available for today's computers. Two such routes are available. One approach is to employ software that is specially developed for coding continuous time data such as that produced in open interaction.<sup>4</sup> The other option is to employ

---

4. Several options are available for real-time coding. One option is Observer XT developed by Noldus Corporation. Another is Mac event Coder and Win event Coder developed by Lisa Troyer ([lltroyer61@gmail.com](mailto:lltroyer61@gmail.com)). Dedoose.com allows the coder to code or prepare transcripts. All have the advantage that the record can be time stamped for each coded event.

video playing software on the computer in conjunction with a word processing program. Time stamping of the video record and repeated coding that is time sensitive are not as easily accomplished with this approach. Whichever route is chosen requires that you work closely with the person(s) you are training to be sure he or she understands the importance of maintaining a secure file of the original data. I have had an assistant inadvertently erase files when trying to code data. As a precaution, I copy the video record to a CD/DVD and have the research assistant work with the copied disk, maintaining a copy of the original file on my office system as well as a video record stored independently of any computer system. As technology advances, these approaches may appear as archaic as tapping pencils to signify acts.

The alternative to coding video records of discussions involves preparing a transcript from the video record and coding interaction from this. The preparation of a transcript can be a time-consuming and enervating process. It also presents an opportunity to engage in some “precoding” by the transcriber. For instance, the transcript may need to identify pauses in interaction of specific length, the interruption of one speaker by another, and the use of back channel comments as the transcript is developed. All of these are time-consuming, often require repeated replays of specific short segments of interaction, and require that the transcriber has a good grasp of your intent. Even if you are not concerned with this level of detail in the transcript, you must still confront how to manage inaudible utterances, the display of overt body language that affects interaction, and environmental noise that may affect the quality of the recording. Time invested in training the transcriber may be as much as you would invest in the actual coders you employ for your analysis. The advantage is that the record can be easily shared with other researchers.

Once the transcript is produced, training a coder is a matter of identifying coding rules such as what the unit of observation is to be, how to classify the activity, and how to recognize regular turn passing as opposed to interruptions. This can often be accomplished by providing the trainee with a set of the coding rules and some instances of coded transcripts you have developed for this purpose. The ideal state of affairs is that you are able to provide this practice coding material from data that are not a part of the actual study material. If this is not possible, it is useful to develop checks to ensure that the work of the assistant is not simply a copy of what you provided. In this case, I try to stress the importance of reliability, but I also make the point that my examples are intended to provide guidance and not to be used as exact models that are to be reproduced. Assessments of reliability with transcript data are often quite high, but there should be some discrepancy between your examples and the data from the assistant.

Assessment of reliability is particularly important in establishing the quality of the training for research assistants and ensuring the integrity of the analysis of the data. Training for coding data should not be done on data you wish to analyze to test hypotheses because it may lead to artificial outcomes. Transcripts from prior projects or transcripts of interaction from groups that do not fit the

scope or initial conditions of the study are appropriate alternatives. The coding of the data that is employed to test the hypotheses should be assessed to ensure that it is appropriately recorded. Various alternatives exist for this, and the selection of appropriate measures depends on the approach taken. The simplest method is to correlate the coding of two or more coders of the same data. More complex measurement models are available and depend on the nature of the coding employed by the investigator.

## C Coding Other Types of Data

Many computer-assisted experiments produce a data file that is ready for analysis or that requires minimal effort to reach this point. In this case, it is often desirable to be sure the new assistant has the necessary skills to translate the data records from one format to another. This may require that you ensure the person has the requisite training to move data from a text file to a spreadsheet and from a spreadsheet to a statistical analysis program. This level of skill is common among undergraduates on most campuses today. If this is not the case with your assistant, these skills can be developed fairly quickly if the person has some level of computer skill. Here again, the raw material makes a difference. The more automation you can design, the better. The possibilities for mistakes in translation are reduced or removed if it is possible to automate the “reencoding” process. If mistakes do occur, they are likely to be much easier to detect because they are more likely to be systematic as opposed to random.

## D Analysis

Once the data files are completed, analysis with statistical software is possible. My experience has been that this is where I am on my own. Most students with whom I have worked have only rudimentary understanding of statistics, and any analysis more complex than a simple one-way analysis of variance is beyond their understanding. There is variation in this expertise, and I have had some undergraduate students work with me who have the skills and experience to carry out complex analyses. It is advisable to develop skills for students by asking them to take up more complex problems.

Careful supervision is in order, however. This is particularly the case if your work requires time series analysis, the testing of stochastic models based outside the scope of the standard linear statistical model, or other more advanced techniques such as structural equation models. Preparing the typical undergraduate to work in this area may be out of the question, but advanced undergraduates and graduate students should be able to work at these more sophisticated levels. At this point, course work in a good statistics program is desirable before the assistant starts working on such analyses. Variations in the levels of expertise that students bring to the project lead me to conclude that enhancing statistical skills of an assistant within the context of a research project is best addressed on an individual basis.

## VI COMPENSATION

Compensation for work in the project may take one of several forms. One of the easiest to grasp is the awarding of academic credit for work on a project. I usually equate hours worked and credit so that the student will be working approximately 3 hours a week for each unit of credit awarded. Financial compensation is governed by local labor market conditions and institutional policy. Navigating this issue is sometimes a daunting process. For instance, I usually pay up to twice the minimum wage for coding and transcription work but would be willing to go higher if the student was exceptionally skilled. Students I have employed to conduct experiments are usually compensated at the same rate. When I have research assistantship available from institutional sources, the rate of pay is fixed by other compensation policies and may include substantially more than this amount. These figures are in a labor market with a scarcity of jobs for students. Intangible compensation is also available and may be more desirable for some students.

A careful plan that specifies how to professionally reward various levels of participation in a study is always desirable before the study begins. Acknowledgment of contributions to a study in a footnote is common in the field. This level of recognition is appropriate for research assistants who have faithfully carried out experiments and contributed to the project by highlighting protocol problems during the collection of data. I usually reserve recognition as a coauthor for individuals who have gone above and beyond the call of duty. Contributions of several types fit this category. For instance, data coding that is particularly onerous and requires the research assistant to participate in the development of a coding scheme and its implementation is the sort of work that may deserve recognition as a coauthor. In other instances, student assistants have suggested analyses that have led to full-blown papers in their own right. In this case, there is no question in my mind that the student deserves credit for this with a role as the coauthor. Finally, if a student were to develop his or her own research project that resulted in a solid study with a publishable paper, I encourage the student to take on the role of lead author in the project and I assume the role as secondary author. A key element in the rubric of how to professionally reward student contributions is that you are open with students about how your plan of recognition is tied to their contributions.

Some prior understanding of the appropriate rules in your field is in order. One of my physicist friends related to me that he is one of 206 authors on an article in his field. The laboratory in which this work took place has a specific set of rules about the inclusion of individuals who oversee its technology and hence 30–40 individuals appear on all of the publications when data originate in this establishment. Others are listed as they contribute to the work. Although sociologists are not likely to encounter this level of complexity in deciding who is listed on a publication, explicit rules are needed. Clarity on this issue will reduce conflict after the fact.

## VII EXPERIMENTAL STAFF AS A GROUP

Thus far, I have addressed training and preparation for research assistants as if there is only one person in the project at any one time. This is often not the case; some projects require a large staff. For instance, experiments may collect data from groups of participants who arrive and participate at the same time, such as a discussion group study or a computer-mediated interaction study in which the appearance of real-time interaction is critical in convincing participants that their interaction is taking place as they participate. In these circumstances, a staff large enough to greet, seat, and conduct postsession interviews with each participant is necessary. The research staff thus could be as large as the groups one is studying. Depending on the availability of staff, you may employ even more individuals due to scheduling problems. Employing a large staff raises significant questions about how to train a number of individuals to work in the study, the roles they will fulfill, and an assurance scheme to be certain that all of them are executing the study in the way you intended when you designed it.

There are several significant issues. For instance, managing the schedules of experimenters can be problematic. Frequently, times of the day, days of the week, and the academic calendar play a role in when participants are likely to be available if you are working on a college or university campus. These concerns typically determine your needs for personnel. I collect data with volunteer participants who receive a nominal payment for their time commitment. My colleagues in psychology employ a subject pool model in which students gain points for their course grades by participating in studies. This difference creates a situation in which they often have difficulty getting participants to sign up and appear early in the academic term; students are engaged in other activities at these times. On the other hand, the cash inducement I pay often ensures a pool early in the term.

At the end of the term, the situation is reversed. The psychology subject pool is full of participants scrambling to improve their course outcomes late in the term, but the interest in earning a few dollars disappears in the face of term papers and reading assignments as final examinations approach. These “facts” of data collection lead to time management concerns for research assistants and the creation of good work environments. Your staff may be as pressed for time at the end of the calendar as your participant pool.

Another issue present in employing a large staff is the dynamic as a social group. Some of your personnel you employ may be “working” for the first time in a cooperative task setting. As a consequence, you may need to engage in socialization about how to work well with others. Lessons in punctuality, the courtesy of informing others if you are ill, how to correct mistakes you make and those made by others, and the necessity to accomplish tasks that are uncomfortable may be necessary for some trainees.

You will also want to determine the extent to which you wish to employ a sink-or-swim strategy as opposed to a strategy that involves careful lessons about

how groups of peers work together to accomplish a task. Whichever model you adopt, my experience is that “getting into the trenches” with the staff is a good idea, at least for a few cycles of data collection. You get a chance to model good behavior in carrying out the study and observe the strengths and weaknesses of members of the staff. This may also give you a chance to communicate the importance of reporting problems to you, as the chief experimenter.<sup>5</sup> I think that it is particularly good strategy to be present during these early phases of data collection. This is especially true if the experiment involves implementing a new protocol. Your presence allows you to catch problems as they occur rather than after several sessions have been completed. It also affords the opportunity to determine if there are problems in the way the project is being carried out by staff, even if they are experienced. The time spent now may be cheap insurance for the entire data collection enterprise.

Perhaps as important as the training that you put in place is the issue of who is in charge. Several different models may be employed, but it is imperative that you, as the senior investigator, communicate how decisions are made, who is to report to whom, the role of protocol in constraining behavior in the laboratory, and your position as the final arbiter of the project. I have worked with students who, once well trained, were provided with the appropriate materials and left in charge of a significant portion of the data collection enterprise, including supervision of a small staff. I have worked with others who required constant supervision on my part, even though they were well trained to begin with. Individual differences in responsible behavior and the ability to work with others frequently determine which model you employ.

Staff members who work together take on a group identity of their own. This can be both helpful and harmful, depending on how it develops and how group members make use of this identity. For instance, some staffs with which I have worked have focused on the task at hand and are thoroughly committed to the success of the project but have no links outside the collection of data. This is certainly a sound example of task cohesion ([Cota, Evans, Dion, Kilik, & Longman, 1995](#)). Other staffs have developed strong social bonds with each other, using “down time” between sessions to play games, compose songs, and otherwise socialize with one another in the laboratory. This may carry over outside the laboratory as well. Staffs of this type are clear examples of groups based on social cohesion ([Aiken, 1992](#); [Cota et al., 1995](#)). Both types have been very successful collecting data and, in my experience, produce high-quality material. The one concern for staffs with high levels of social cohesion is that monitoring may need to be more intentional to ensure the group stays on task. Deliberate efforts to ensure one of these outcomes are probably not likely to be successful.

---

5. Students often think of such reporting as “ratting” on their peers. You must communicate the importance of good information in these situations, even if there are unpleasant consequences for members of the staff. The collection of experimental data is costly; errors make it more costly.

My experience is that interaction dynamics within the staff lead to one outcome or the other independent of my actions.

### **A Communicating with a Staff**

You may find that periodic meetings with staff are valuable for reinforcing training, communicating about problems encountered in collecting data, and learning about how effectively the staff is working together. The timing of such meetings and their content are variables under your control, except in emergency situations. A simple rule of thumb is that such meetings should be planned in advance, notice given to the staff so they can attend, and a productive agenda created. Meeting simply for the sake of meeting is not likely to endear you to your experimenters. That said, it may be useful to schedule meetings frequently early in a data collection sequence and use them as training and reinforcement events and then taper off as the collection process hits its stride. Later meetings may be designed to catch up on progress in completing the data collection design, identification of issues of suspicion, lack of satisfying scope or initial conditions in the study protocol, or other problems in the collection of data.

Telephone calls, e-mails, and text messages may be useful from time to time, but I am always concerned about the tendency for these “conversations” to become monologues on my part about the experiment. I prefer face-to-face meetings when issues of substantive significance arise in the data collection process. Learning that a particular day and time are experiencing large numbers of “no shows” is best handled with these communication techniques, but brainstorming how to solve a general no-show problem may not be.

Encouraging staff to talk in these meetings is critical. Good experimenters will identify problems in the protocol. You want to know about these before they create significant artifacts in your data that imperil the validity of your study. Develop a strategy to encourage participation in these discussions as peers so that all of the members of the group contribute. Going around the table asking for contributions may work, but you may need to do more to break down status differences, especially if your student help views you as the “boss.” An icebreaker can be helpful in initiating these conversations. Often, a tale of a missed opportunity to ask a question in a project or a tale of the experiment that was a complete mess creates the climate to encourage candid and informative discussion.

## **VIII TRAINING FOR POSTSESSION INTERVIEWS**

In addition to translation of computer-generated records, you may ask your assistant to collect and code data from postsession interviews. This step is particularly critical when deciding which data to employ in testing your hypotheses, and it requires as much care in training as does coding open interaction. The objective for the research project is to identify those subjects for inclusion in the analysis who meet the scope conditions of the theory and who also satisfy

the initial conditions for their assignment to a particular condition of the theory. For instance, in expectation states research, individuals must be task focused and collectively oriented as well as believe that they are better than, equal to, or worse than their partners in order to include the data in the analysis.

Listening to audio recordings of interviews with the trainee may provide an important starting point for this training. Working with the assistant to identify key terms, verbal cues such as tone of voice, and other paraverbal markers such as long pauses may provide valuable information for both you and the trainee. Learning to identify body language in the interview may require that you video record some interviews or group discussions to provide examples for your trainee. Generally, this is time well spent, as is the careful development of a protocol that identifies appropriate decision rules for excluding data from a particular subject. It may also be advisable to provide a set of guidelines about when to consult senior investigators about including or excluding data when an interview produces ambiguous information.

Supervised practice is always advisable until the trainee develops proficiency in conducting the interview. Periodic checking is advisable once the person is working by himself or herself. I have begun to review all tapes of postsession interviews to determine if the appropriate questions have been asked and the appropriate decisions made by the interviewer. This can be an arduous and tedious task if the project is a large one, and it may be desirable to develop a sampling regimen if this is the case. It may be advisable to carry out this decision-making as part of group meetings with the staff. Employing multiple decision makers will help increase the reliability of these choices. If the project has a large number of cases, sampling may be the most expedient way to maintain a consistent set of rules among staff members.

At the very least, the assistant should be able to identify whether or not an individual participant satisfies the scope and initial conditions for inclusion in a data set for testing hypotheses. This is often a subtle and complex decision-making process. Subjects may meet scope conditions but not initial conditions, for instance. In this case, the individual should be excluded from the analysis. It is often valuable when training a new assistant to ask him or her to determine whether this outcome is unique to this subject or is part of a pattern in the data collection that requires systematic attention and perhaps intervention in the protocol. At the time I wrote the first version of this chapter, I was pretesting a new experiment that was not successful in its implementation. I encouraged the assistant, who was in training for this study, to identify information from the interviews to assist us in revising the protocol. This proved particularly valuable as he began to develop a deeper understanding of the project. He also became much more sensitive to the success and failure of our attempts to manipulate feelings of our subjects toward one another and their orientation to the collective task.

These end-of-session data are also valuable to you when assessing the relative capabilities of various assistants as well as the success of the experimental

protocol. If the project is large enough to warrant several assistants, comparison of rates of rejection of participants from the analysis pool may be in order. These comparisons allow you to identify experimental conditions that are not accomplishing the intended effects if rejection is consistently high across several experimenters. Alternatively, an experimenter whose rejection rate is inordinately high in comparison to rates of other experimenters may require special attention to determine if the person is applying rules too stringently. The complementary situation may apply if the assistant is not rejecting as many as his or her fellows. This monitoring is valuable because it may signal a variety of problems with decision-making by the research staff. Each of these comparisons provides important information about the success of the experiment.

## **IX ESTABLISHING A MENTOR ROLE**

In the introduction, I mentioned wanting to produce research staff as passionate about my research program as I am. This does not happen often, but it is possible to create conditions that lead students to an appreciation of the value of social science experimental work. This often is possible if a mentor–mentee relationship can be established with individuals on your research staff. Such a relationship requires an intentional effort on your part to create a task ladder of increasing skill development. Associating it with increasing responsibility in the research program is necessary. I must confess that I have not had much success creating research assistants who go on to independent careers in experimental social science, due in part to the nature of the student population at my institution. Most of these students are focused on careers in the criminal justice system.

Simple ideas about how to create a task ladder to increase skills are best. Such a series of tasks may begin with the “easiest” script, the shortest time commitment, or the simplest interviews. Whatever approach one takes, it should be apparent to the person who is learning in this situation that there is a sequence of skill development and responsibility to be mastered. As one is successful at one level, advancement is possible. One marker of this may be the written report.

I ask my assistants to keep a log of their experience with a particular subject. This log includes information about the conduct of the experimental session, the interview, notes about how the decision was reached to keep or reject data of a particular participant, and a summary of the session. This is very informal note taking, but it serves as a record for the construction of other written reports. The mechanism of keeping this log is open. It is possible to store this information in computer files. I have used large catalog envelopes for each experimental session in the past, but these begin to present storage problems and their discrete character and ease of separation defeat the purpose of using them as the data for descriptive and analytic material. I have become a convert to hard copy in a three-ring binder. This approach is antiquated and giving way to the technological solution of a word processor.

The objective is to have all of the information in one “physical” place, as well to allow investigators to rearrange “pages” as needed to prepare reports. For instance, it may be desirable to review data from all of the rejected participants at the same time, data of all accepted participants, data from only one condition, data from a range of dates, or data collected by a particular assistant. The ease of manipulation of integrated records makes this relatively simple.

Recent developments suggest that using word processing files, synced with the actual data record of the experiment, may be more valuable. These files are easily searched for common themes, easily analyzed for testing theories, and may be easily archived for sharing with other investigators. One must be cautious about data integrity, however. One hard drive crash and the record of the entire experiment may be lost. This possibility should encourage storage of the data in multiple sites with a variety of media in use to ensure recovery if some sites or media are lost.

The point of keeping detailed records of experimental sessions becomes apparent to the trainees when you ask them to write reports. These reports may be of several sorts. One possibility is a report for the other members of the research staff about what is going well and what is not going well as the experimental sessions proceed. The frequency of preparing such reports will depend on the rate at which data are being collected and your interest in being able to monitor progress. It may be desirable to have a report each week or at the conclusion of particular phases of data collection, such as after the first 50 acceptable cases, the first month of data collection, or the completion of a particular set of experiments.

A key issue in this report preparation is the audience that is being addressed. Initially, it is appropriate to have reports prepared for internal audiences only. My preference is to ask that the first report be prepared for me. This gives the new research assistant a chance to write for a known audience. It can also be a nonthreatening experience if you present the task as one of providing information about how the process of data collection is proceeding. The expectation is that the report will provide valuable information about progress, but it may be written in an informal prose style.

Subsequent reports may be prepared with different audiences in mind: colleagues in the same research area who are interested in the progress of your work; colleagues in other fields interested in the work but relatively uninformed about the approach to data collection you are undertaking; and, finally, a section of a research report intended for professional presentation in a conference paper or research article. The key to this process is to stress the developmental nature of the report-writing process. I usually do this by relating the fact that one of my publications went through something on the order of 20 drafts before it saw print. This tends to reduce the anxiety of the new research assistant and also make the point that even seasoned professionals do not “get it right” the first time.

A matter of pure judgment must be addressed at this point. It is always desirable to consider when to raise the bar for the research assistant as he or she acquires new skills and develops maturity. Part of the answer to this question is an easy one. Once the person has developed appropriate skills and is confident in their application, it may be time to ask him or her to learn new, more complex skills. On the other hand, issues of maturity and motivation play an important role in decisions about increasing responsibility and the appropriate level to identify for further advancement. Some of my assistants have mastered all of the skills I have asked them to but their interest in going further is not present. Others have wanted to take on more responsibility but have not demonstrated the maturity to do so.

## X CONCLUDING REMARKS

A good experiment is like a good play: sound script, excellent staging, good use of props, and a superb cast. There are some differences: the script in the experiment is open ended. No one, including the researcher, knows precisely how it will come out in the end. We may know how we want things to come out, but the actual outcomes depend on our ability to stage a good production, have the cast put on a convincing performance, and create the suspension of disbelief critical to the production of high-quality data. An experimenter also encourages a strict, repetitious performance of the “play.” Idiosyncratic actions, improvised lines, and variation in emphases may be valued in the theater, but they are cause for substantial concern in an experiment in which one is trying to create consistent social conditions and the only variation is that created and authorized by the experimenter.

Every so often, some of these elements fail. If the script is not doing what we want, we may revise it after the pretest. In many ways, this is the easiest element to adjust. If the staging is wrong (lighting is bad and the computers fail too often), we can effect repairs. Sometimes it is necessary to remove a research assistant who is not performing satisfactorily. This is always unpleasant, but it may be necessary. If the person is routinely late or misses assigned work times regularly, this may be a message that he or she wants to leave the study but does not wish to quit. Dismissal is in order. Sometimes the product will be poor in that data from sessions executed by an individual turn up bad too often. Again, dismissal or reassignment may be in order. This is always a difficult task, but if you can have a conversation specifying the problem and how to solve the problem, it may prove valuable to both you and the assistant who is being released.

Having said all of that, I have not dismissed many assistants during a 30+-year career of experimental work. I suspect that requiring the classroom experience in a research methods course may account for some of this stability. Once students have been through the methods course, they typically understand on some level the data collection process. The training and work you provide are a practical experience that extends their classroom learning.

Good students will appreciate this. Those unlikely to appreciate the experience generally do not apply. Training those who do is an exciting and valuable enterprise. Success often leads to collaborators in your research program.

## ACKNOWLEDGMENTS

I acknowledge the contributions of individuals who have shared their experiences and laboratory resources with me. Hans Lee, my dissertation advisor, provided valuable experience in how to train personnel for a laboratory setting and emphasized the value of keeping carefully prepared laboratory manuals. David Wagner of the University at Albany provided his laboratory manual so I was not working only from my own memories and materials. Alison Bianchi of the University of Iowa graciously read parts of the manuscript. My wife, Ann Converse Shelly, provided valuable feedback as the manuscript was developed. Finally, I thank Murray Hudson, Professor Emeritus of Theater at Ashland University, and Laura Bares Shelly for their generous sharing of ideas about theatrical production and the role of the director and property manager in developing a performance that depends for its success on the suspension of disbelief.

## REFERENCES

- Aiken, L. R. (1992). Some measures of interpersonal attraction and group cohesiveness. *Educational and Psychological Measurement*, 52, 63–67.
- Bales, R. F. (1950). *Interaction process analysis*. New York: Addison-Wesley.
- Cohen, B. P. (1989). *Developing sociological knowledge: Theory and method*. Chicago: Nelson Hall.
- Cota, A. A., Evans, C. R., Dion, K. L., Kilik, L., & Longman, R. S. (1995). The structure of group cohesion. *Personality and Social Psychology Bulletin*, 21, 572–580.
- Rosenthal, R. (1976). *Experimenter effects in behavioral research*. New York: Irvington.
- Rosnow, R. L., & Rosenthal, R. (1997). *People studying people: Artifacts and ethics in behavioral research*. New York: Freeman.
- Willis, J. R. (1994). *Directing in the theatre: A casebook* (2nd ed). Metuchen, NJ: Scarecrow Press.

## Chapter 5

# Human Participants in Laboratory Experiments in the Social Sciences

**Will Kalkhoff**

*Kent State University, Kent, Ohio*

**Reef Youngreen**

*University of Massachusetts–Boston, Boston, Massachusetts*

**Leda Nath**

*University of Wisconsin–Whitewater, Whitewater, Wisconsin*

**Michael J. Lovaglia**

*University of Iowa, Iowa City, Iowa*

## I INTRODUCTION

Laboratory experiments in the social sciences are concentrated in four disciplines: psychology, sociology, political science, and economics. We focus on the recruitment of participants in the first three. (Davis and Holt (1993) and Kagel and Roth (1995) provide extensive details of experimental methods in economics.) Experiments using human participants represent a relatively small proportion of the research in each of these disciplines. For example, much of psychology is not social, focusing instead on individual differences unrelated to social interaction or on the social behavior of nonhuman animals. The subdiscipline of social psychology is primarily experimental in psychology but spills across the disciplinary boundary with sociology. In sociological social psychology, experiments have had a place for several decades, but in uneasy co-existence with less intrusive observational methods. In sociology, experiments are concentrated in a subdiscipline known as “group processes.” Compared to psychology and sociology, laboratory experiments in political science have only recently become influential in the direction of research in the field. In all of these disciplines, however, the influence of experimental research on the field has been disproportionately large relative to its volume.

Szmatka and Lovaglia (1996) explain the peculiar “convincingness” of the results of laboratory experiments compared to results from other research methods as a combination of reproducibility, incremental adjustment of research design to counter criticism, and transparency. A critic can be invited into the laboratory to see for him- or herself with relatively little investment of time and resources. (See also Lovaglia (2003) on the power of experiments.)

Methodological issues are often couched in theoretical and even moral terms as practitioners of a particular technique seek to carve out a resource-rich niche for their work (Szmatka & Lovaglia, 1996). The treatment of participants in experiments is no exception. Whether participants ought to be volunteers, paid for their efforts, or required to participate to complete a class or degree has been a hotly contested issue that sociological experimenters have used to distinguish their research from that of psychologists. Sociological researchers have argued for greater care in avoiding coercion to participate, while psychological researchers have maintained that participation in experiments is a vital part of undergraduate education and thus should be required.

Within sociology, the debate over the ethics of deceiving participants has polarized experimental and nonexperimental social psychologists. Should “informed consent” require that all the relevant issues involved in the research be explained to participants before they agree to continue with it? On the one hand, experimenters using deception note that few important topics could be effectively researched in the lab without deception. For example, explaining in advance that the study investigates racist tendencies of participants would certainly alter their behavior during the study, thereby masking any relevant effect. On the other hand, those decrying deception maintain that the damage to the reputation of the discipline caused by deceiving participants far outweighs any contribution to knowledge that such techniques produce. Experimenters counter that participants should expect the reality they encounter in an experiment to match the reality outside the lab no more than they expect the same congruence from a theatrical production. Ironically, opponents of constructing alternate realities in the lab often come from a social constructivist ideological position that also denigrates experimenters as “positivist,” often using the term in straw-man fashion to misrepresent scientific activity (for related arguments, see Turner, 2006). In practice, the issue has been resolved by institutional review boards (IRBs) that approve experimental designs. IRBs have consistently upheld the value and ethical soundness of experiments using deception when approved procedures are followed and participants thoroughly debriefed after the experiment.

However, as Hertwig and Ortmann (2008) have suggested, upholding the rights of research participants and ensuring their welfare is arguably not the only ethical matter at stake when it comes to deception. The spirit of their argument is “Fool me once, won’t be fooled again.” A participant once deceived in a

study (or who knows someone once deceived) is likely to *expect* to be deceived in other studies, which undermines the foundation of experimental research. Although the potential for this problem has led to a general prohibition against the use of deception in economics (Hertwig & Ortmann, 2008), a less extreme alternative is to point out that researchers have an ethical obligation to take seriously how the use of deception may impact not only participant welfare but also the integrity of a study and its findings.

Although theoretical and moral debates about participants are interesting and important, the main purpose of this chapter is to describe the ways that participants are recruited for laboratory experiments in enough detail to allow a researcher setting up a laboratory to efficiently recruit participants.

## II HUMAN PARTICIPANTS IN PSYCHOLOGY

Psychology has been concerned with the problem of “recruiting participants” for decades. The uncertainty (How do we get participants?) faced by other disciplines that have more recently come to utilize laboratory experimentation has encouraged a “mimetic process” (DiMaggio & Powell, 1983) whereby the newer players have tended to model psychology’s tried and true procedures. In this section, we begin by describing these procedures in terms of what is most used throughout the country. We then turn to a discussion of important ethical and methodological issues that present themselves in relation to methods for recruiting participants. Next, we discuss some relatively recent Web-based technological innovations for managing participants once they have been recruited. We close with a summary of recommendations for developing procedures for recruiting participants and maintaining participant pools based on lessons from psychology.

### A How Are Participants Solicited?

Two major surveys of participant pool practices in psychology include specific questions concerning how research participants are “solicited” (i.e., recruited). More than 30 years ago, Miller (1981) surveyed the top 100 most cited universities in the United States, Canada, and the United Kingdom based on total citations in the Science Citation Index for 1978. A total of 76 universities responded (although no responses were received from the United Kingdom), 70 of which included some form of introductory psychology participant pool. Of these, 8.6% relied on *volunteering* as the basis of participation, 12.9% gave *extra credit* for experimental participation, 14.3% offered *extra credit with other options* for earning the same credit, 2.9% *required participation*, 58.6% *required participation but offered options* for fulfillment of the requirement, and 2.9% used some *other means* of recruiting participants.

In the late 1990s, Landrum and Chastain (1999) reused Miller’s (1981) instrument in an exhaustive survey of psychology departments without graduate

programs ( $N=1238$ ).<sup>1</sup> Of the 570 universities that responded (47.6%), 478 provided “usable responses” (Landrum & Chastain, 1999, p. 31). Their results indicate that 34.7% of the departments in their sample currently rely on *volunteerism* as the basis of participation, 18.9% give *extra credit* for experimental participation, 24.8% offer *extra credit with other options* for earning the same credit, 3.3% *require participation*, 14.1% *require participation but offer options* for fulfillment of the requirement, and 4.2% employ some *other means* of recruiting participants.

A comparison of Miller’s (1981) and Landrum and Chastain’s (1999) results suggests that psychology departments had come to rely more on volunteerism over time. Psychology departments today may also be less likely to require participation (with options) than in previous years. The apparent increasing reliance on volunteerism makes sense given the proliferation of federal and professional association rules governing the involvement of human participants in research along with university IRBs operating to ensure researcher compliance with these rules (discussed later). By that argument, however, it is surprising that so many departments in the not-so-distant past were in discord with federal and professional association rules by requiring participation without offering options to fulfill the requirement. Admittedly, however, the differences between Miller’s sample of elite institutions and Landrum and Chastain’s sample of undergraduate-only psychology departments complicates interpretation of the observed trends, and it is likely that we would reach different conclusions about trends if Miller’s study were replicated. That being said, we turn now from a description of what departments have been doing to a discussion of what perhaps they *ought* to be doing.

## B Ethical Considerations

In developing ways to recruit participants, set up participant pools, and carry out experiments, researchers have confronted two main ethical concerns: (1) how to avoid coercive practices; and (2) how to ensure that research participation has educational value.

### 1 Coerciveness

As indicated previously, approximately one-third of psychology departments relied on the use of “totally volunteer” participants for research participation as of the late 1990s (Landrum & Chastain, 1999), and that number has likely increased since then. On the other end of the spectrum, slightly more than 3%

---

1. Landrum and Chastain (1999) offer the following explanation as to why they surveyed departments with only an undergraduate program in psychology: “[T]here seems to be some indication of a general increase in research productivity at undergraduate institutions (based on informal conversations with colleagues at conferences) and because our own university has a psychology department without a graduate program” (p. 30).

of psychology departments required experimental participation of its students as we moved into the 2000s. Again, requiring participation (without options) is problematic because federal rules, as well as guidelines of both the American Psychological Association and the American Sociological Association, state that individuals should not in any way be coerced to participate in research. Accordingly, some departments (~14% by the late 1990s) offer other options to fulfill research requirements ([Landrum & Chastain, 1999](#)). Traditionally, these involve having students who “opt out” to write a paper, complete additional coursework, or take a quiz ([Sieber & Saks, 1989](#)). However, although the availability of such alternatives may seem to minimize coerciveness (i.e., insofar as *any* alternative is presented), a student faced with choosing between participating in an experiment and completing a potentially dull and more time-consuming alternative is actually faced with a “Hobson’s choice” (i.e., an apparently free choice that is not really a choice). In other words, “Limited and unattractive alternatives [still] constitute an element of coercion” ([Sieber & Saks, 1989, p. 1057](#)). However, [Trafimow, Madson, and Gwizdowski \(2006\)](#) found that although more than half of the undergraduate participants in their investigation reported perceiving that an alternative writing assignment would (probably) require more effort than participating in an experiment, the majority also indicated they would choose experimental participation over writing a paper when the two were described as being “equally effortful and time-consuming” (p. 248). This may suggest that experimental participation is at least somewhat attractive to undergraduates, all else being equal. Recent research confirms this. [Roberts and Allen \(2013\)](#) developed a new brief measure of student perceptions of the educational value of research participation and found that university students in their sample saw more benefits to research participation than costs. However, to limit coercion, we share the view that researchers must think conscientiously and creatively about developing meaningful alternatives to required participation in experiments (for a list of suggestions, see [McCord, 1991](#)).

As another kind of work-around to the problem of coerciveness, many psychology departments do not require student participation in experiments but instead offer some kind of extra credit in exchange for participation ([Landrum & Chastain, 1999](#)). As of the late 1990s, some did so without offering other options to earn equivalent credit. Again, students faced with this situation may believe they have no choice but to participate in experiments (i.e., if they wish to improve their grades). [Leak \(1981\)](#) showed that students in one program found this kind of extra credit “a temptation somewhat hard to refuse” (p. 148). The bottom line is that it is a mistake for researchers to assume that coerciveness is avoided by offering extra credit for participation instead of making participation a requirement. The situation is somewhat improved when departments offer extra credit with *options to fulfill the requirement*. Here again, however, researchers must think carefully about developing *meaningful* extra credit options instead of going no further than just presenting some kind of option that, although perhaps easy to evaluate and tally (e.g., a paper), is unattractive to students.

## 2 Educational Value

In the early 1980s, [Miller \(1981\)](#) noted that “little has been done to provide evidence that experimental participation is a valuable pedagogic device” (p. 213). An exception noted by Miller was [Britton \(1979\)](#), who found that students required to participate in psychology experiments at a large southern state university rated the educational value of the requirement “considerably below the maximum” (p. 197). Miller’s observation that scant evidence existed regarding the educational aspects of experimental participation, along with Britton’s discouraging finding, constituted something of an embarrassment because psychology departments had long before then justified such participation mainly in terms of its ostensive educational value ([Jung, 1969](#)). Consequently, Miller’s work inspired a number of studies dealing with the issue of the potential educational value of experimental participation. With a few exceptions (e.g., [Coulter, 1986](#); but for a critique of Coulter’s research, see [Daniel, 1987](#)), the weight of the evidence accrued during approximately the past 30 years suggests that research participation does have educational value. For instance, [Rosell et al. \(2005\)](#) followed psychology undergraduates during the course of a semester and reported that their improved understanding of research methods may have been due to their participation in research. Similarly, [Gil-Gómez de Liano, León, and Pascual-Ezama \(2012\)](#) provide evidence that research participation improves students’ exam performance. In addition, focusing specifically on emotional responses to research participation along with educational value, [Flagel, Best, and Hunter \(2007\)](#) surveyed 101 research volunteers and found that the majority self-reported no stress and perceived their participation as a valuable learning experience.

That said, some methods of soliciting participants clearly prove better than others with regard to enhancing the pedagogical value of experimental participation. As mentioned previously, [Britton \(1979\)](#) found that students in the department he investigated did not rate their experimental participation favorably in terms of its educational value, but these students were required to participate in experimental research. Indeed, coercive policies and student perceptions of educational value appear to be negatively related. [Nimmer and Handelsman \(1992\)](#) present results from a quasi-experiment showing that students reported more positive attitudes toward the learning value of research participation when they were exposed to a “semivoluntary” research requirement as opposed to a mandatory research requirement (i.e., they could participate in research in exchange for extra credit toward their course grade). However, although offering extra credit as an incentive for research participation may produce more favorable student perceptions of educational value vis-à-vis mandatory participation, alone it does not seem to go a long way toward meeting learning objectives. [Padilla-Walker, Zamboanga, Thompson, and Schmersal \(2005\)](#) found that among 193 students enrolled in an introductory psychology class at Midwestern State University, those who opted to participate in research for extra credit were already high academic achievers. Those who ostensibly had the most to gain

from the experience were the least likely to participate. Specifically, of those who took advantage of the extra credit research opportunity, 70% were students with excellent or good grades, whereas only 3% were students with below average grades.

The most promising results seem to come from studies examining psychology programs that require participation but also offer options to fulfilling the requirement. [Landrum and Chastain \(1995\)](#) examined the educational value of research participation among students enrolled in a general psychology course at Boise State University in which each student had to “complete some sort of outside-of-class *activity* exposing him or her to psychological research” (p. 5, italics added). Most students opted to be research participants, but in the context of being able to choose between experimental participation and a thoughtful alternative, they tended to “agree” that their participation: (1) helped them to learn about psychology; (2) helped them to understand research better; and (3) added variety to the course. Furthermore, these students tended to disagree that their research experience was a “waste of time.”

However, when faced with less meaningful alternatives to participating in research (e.g., writing a paper), the educational benefits of a research requirement may not be as robust. In a study of 774 general psychology students enrolled at Central Connecticut State University during a 4-year span, [Bowman and Waite \(2003\)](#) found that students who chose to participate in an actual research study were more satisfied with their experiences than those who wrote an optional paper. This finding underscores the key point we made previously: researchers should think carefully about devising meaningful alternatives to required participation in research. Doing so appears to both reduce coerciveness and enhance student learning.

However, the best way to know whether one’s own departmental policies effectively address these ethical concerns is, of course, to check and see. Toward doing so, [Landrum and Chastain \(1995\)](#) describe a short, easy-to-administer “spot-check” form developed to assess the educational value of undergraduate participation in research. [Leak \(1981\)](#) describes a brief 10-item post-study questionnaire that can be used to gather information on students’ perceptions of both coercion and educational value in relation to their participation in research activities. Also, as mentioned previously, [Roberts and Allen \(2013\)](#) have developed a new brief measure of student perceptions of the educational value of research participation. Researchers in departments with experimental programs will find these useful for performing “quality control” checks of their practices for carrying out research.

## C Methodological Considerations

The most common criticism of laboratory experiments is that their findings lack external validity. Not surprisingly, then, methodological issues in relation to participant recruitment practices and participant pool policies tend to revolve

around how the unique characteristics of college-student research volunteers threaten the generalizability of such data (see, e.g., Rosenthal & Rosnow, 1969). However, if the goal of laboratory experiments is *not* to statistically generalize findings, then such criticisms constitute a “red herring” insofar as they draw critical attention away from other truly relevant (and potentially serious) methodological problems.

We share the view that the purpose of laboratory experiments is to *test theories*. In the sense that we are using the term, a “theory” is a set of logically related propositions specifying expected cause-and-effect relationships among variables under specific “scope conditions.” Scope conditions, the lynchpin of scientific theory testing, are statements about a theory’s domain of applicability. For example, status characteristics theory (e.g., Berger, Fisek, Norman, & Zelditch, 1977) predicts that a group’s higher-status members will be more influential than its lower-status members, but only when the group is *task oriented* and *collectively oriented*. We expect the claims of status characteristics theory to hold true whenever and wherever its scope conditions are met. Therefore, as Lucas (2003) states:

*Criticizing an experimental test of status characteristics theory that employs undergraduate students as having low external validity because results cannot be generalized to a larger population misses the point. The theory makes propositions on human behavior unbounded by the particulars of population parameters. (p. 241)*

With such an understanding, one can begin to appreciate the *advantages* of employing homogeneous nonprobability samples (e.g., college-student volunteers) in the service of theory testing. The main advantage is summarized by Calder, Phillips, and Tybout (1981):

*Homogeneous respondents ... are preferred because they decrease the chance of making a false conclusion about whether there is covariation between the variables under study. When respondents are heterogeneous with respect to characteristics that affect their responses, the error variance is increased and the sensitivity of statistical tests in identifying the significant relationships declines. Thus, heterogeneous respondents ... increase the chance of making a Type II error and concluding that a theory was disconfirmed when, in fact, the theoretical relationship existed but was obscured by variability in the data attributable to nontheoretical constructs. (p. 200)*

From this point of view, we are encouraged to think about participant recruitment practices as a potential source of *measurement error* (Thye, 2000). Thus, rather than attempt to develop participant recruitment practices that increase heterogeneity and maximize the generalizability of findings (as conventional wisdom might advise), the researcher seeking to conduct a laboratory test of a well-specified theory should strive to do just the opposite. The key methodological issue concerns how participant recruitment practices can be tailored to *increase* (rather than decrease) the homogeneity of participant pools.

To illustrate the problem, let us consider a fable describing the usual exigencies faced by experimental psychologists in relation to recruiting participants:

*The interior decoration of psychology departments is done on an ad hoc basis, but the departments share a remarkable commonality. Toward the beginning of each year bulletin boards contain a few notes about joining clubs, renting spare rooms, and cheap prices for yesterday's computers. After the first few weeks of the term, new notices go up asking for volunteer subjects saying it only takes 10 minutes. You remember these notices; they attract the "good of science" volunteers. A few weeks later, because these good-hearted volunteers are a scarce resource, the notices start offering either credits for your courses or monetary payment. (Wright, 1998, p. 99)*

If we regard participant recruitment practices as a possible source of measurement error, the problem as illustrated by the preceding fable is that such practices are often treated in an arbitrary way. The experimentalist who makes a practice of doing whatever he or she has to do at a given point in the semester to amass research volunteers will likely increase the heterogeneity of the sample. He or she will probably recruit different kinds of volunteers at different times, depending on the particular recruitment method used to build up the participant pool. As we are suggesting, one result at the study level is increased variability in measurements and consequently, as [Calder et al. \(1981\)](#) note, an increase in the likelihood of Type II statistical error (i.e., false acceptance of the null hypothesis). Another potential problem is a breakdown of random assignment insofar as personality characteristics and demographic (e.g., sex and employment) and academic factors (e.g., grade point average and year in school) exogenously contribute to determining *when* people participate in a study during the course of a semester ([Aviv, Zelenski, Rallo, & Larsen, 2002](#); [Bernard, 2000](#); [Stevens & Ash, 2001](#); [Witt, Donnellan, & Orlando, 2011](#)). In summary, although not widely recognized, participant recruitment methods and self-selection have important implications for theory testing. By employing recruitment procedures uniformly and by employing strategies to address individual differences that affect the timing of research participation (for specific recommendations, see [Aviv et al., 2002](#)), the researcher ensures that the participant pool will be as homogenous as possible and that truer random assignment is achieved. Consequently, "the researcher can be more confident that any negative results reflect failure of the theoretical explanation" ([Calder et al., 1981, p. 200](#)).

## D Managing a Participant Pool

Regardless of how research participants are solicited—whether by volunteerism, requirement, offering extra credit, or some other method—they of course still have to be scheduled for experimental sessions once they decide to participate. In the past, most psychology departments used sign-up sheets posted on public bulletin boards to schedule student appointments for participation in research.

This method is problematic, however, because federal regulations concerning the privacy rights of research participants speak against it. Protections for participants mandate that students should not have access to a list of who else has signed up for studies.

As an alternative to paper-and-pencil sign-up sheets, researchers are increasingly using automated systems administered via the Web to manage human participant pools. For example, one of the more commonly used online programs, that by Sona Systems Ltd. (<http://www.sona-systems.com>), works as follows. After being informed in their classes about existing research requirements, volunteer opportunities, or the chance to earn extra credit in exchange for research participation, students are directed to a secure Web site where they can read more about their department's research policies and procedures, create a personal account, fill out any prescreening surveys (if desired), view a listing of studies currently available to them, and schedule an appointment online by selecting among vacant time slots posted by researchers. Students receive appointment confirmations and reminders sent to an e-mail account of their choosing, and they can cancel their appointments up to a certain point if allowed to do so by the researcher. After an experimental session, the researcher uses the Web interface to document a student's participation and, if appropriate, award credit. Most important, this information can be retained in a database to prevent students from signing up for a given experiment more than once. Finally, in cases in which credit is awarded, instructors can access the Web system through a separate password-protected entry point to retrieve information regarding their students' participation in research (which studies they completed, how much credit they earned, etc.).

Overall, Web-based experiment management systems are quite an improvement over sign-up sheets. First, by allowing students to create their own private accounts on secure Web servers, Web-based systems are more compliant with government protections concerning individuals' privacy rights. Second, the automated systems save time and money and reduce errors associated with manual processes. Researchers no longer have to spend their time preparing sign-up sheets and worrying about paperwork errors, nor do they have to set aside funds to pay phone schedulers and cover the costs of photocopying sign-up sheets and reminder slips. Third, Web-based experiment management systems are convenient for participants, instructors, and researchers alike. Participants can sign up for experiments (or cancel appointments if given the option) from any place they have Web access. Instructors can log on to the system from their offices or elsewhere to generate reports indicating which of their students have been awarded credit for participation. Site administrators and research personnel can access the system through their own password-protected entry points in order to edit studies and time slots and to find out which sessions in the schedule have been filled, canceled, etc. The latter feature is particularly useful because most systems allow administrators to prevent students from signing up too late for experimental sessions. That is, most systems can be set up to "close" an

experimental session if it is not filled by a certain point (e.g., 12 hours in advance). Experimenters can check the system remotely and be certain of their appointments, or lack thereof, well in advance. A fourth important general benefit of Web-based experiment management systems is that they work to enhance experimental/methodological integrity in a variety of ways. As mentioned previously, most systems include automated checks preventing students from signing up for an experiment more than once. Many online programs (e.g., Sona) also allow researchers to post “prescreening surveys” that determine each student’s eligibility for experiments and automatically provide the student with a customized listing of available studies. Furthermore, with the convenience they afford, and by sending appointment confirmations and reminders via e-mail, the newer Web-based experiment management systems may increase overall participation rates by as much as 50% and reduce the rate of “no-shows” to 5% or even less ([Sona Systems, 2013](#)).

One downside is cost. Annual fees for programs specifically designed for experiment management can exceed \$1,000. However, where appointment scheduling and reminders are the primary concern, and where advanced features such as prescreening surveys, participant tracking, and account creation are not essential, other more reasonably priced or even free online calendar tools are an option. For example, YouCanBook.Me (<http://gb.youcanbook.me>) can take information from a researcher’s Google Calendar and show available appointments to participants in an online, bookable calendar. The standard version is free, and more sophisticated versions that afford more features (e.g., the ability to link to multiple calendars and send e-mail or even SMS/text message reminders) are available at comparably low monthly rates.

## E Lessons from Psychology: A Summary

The task of developing procedures for recruiting participants and managing participant pools, even though it may seem easy on the surface, can be quite daunting when a researcher comes face-to-face with all of the ethical, methodological, and technical issues at stake. Based on our review of psychology’s contributions to addressing these important issues, we close by offering a summary of six important recommendations to serve as a primer for those new to the process of building an ethically and functionally sound infrastructure to support laboratory research. Researchers seeking to enhance their existing procedures for recruiting participants and managing participant pools may also find the following suggestions useful. Our recommendations are listed in [Table 5.1](#).

## III PARTICIPANTS IN SOCIOLOGY

As mentioned previously, laboratory experiments in sociology have a long history. [Jackson and Cox \(2013\)](#) offer an excellent introduction to the basic principles of experimental design, illustrate the relevance of experiments to sociology,

**TABLE 5.1** Summary of Recommendations for Developing Procedures for Recruiting Participants and Managing Participant Pools

Recommendation	Details
1. Incorporate what works.	<a href="#">Sieber and Saks (1989)</a> provide copious details on a set of “exemplary subject pool documents.” Researchers will find helpful the guidelines that Sieber and Saks propose for participant pool documents concerning: (1) instructions for student participants; (2) instructions for participant pool users; and (3) the administration of participant pools (see pp. 1058–1061).
2. Develop ethically responsible policies and procedures for recruiting participants.	In line with federal and professional association rules, researchers must ensure participant privacy and avoid coercive participant recruitment practices. Privacy issues can be addressed by using Web-based experiment management systems as opposed to public sign-up sheets. Coerciveness is minimized when researchers provide meaningful alternatives to research participation, such as attending research colloquia or serving as observers in studies (for more suggestions, see <a href="#">McCord, 1991</a> ). Offering these kinds of meaningful alternatives should also serve to enhance the educational value of such research experiences.
3. Get your policies and procedures approved.	Once departmental policies and procedures for recruiting participants and participant-pool maintenance have been developed, they must be approved by your institution’s IRB prior to being implemented. Also, researchers can get answers to compliance-related questions from their respective IRBs.
4. Implement quality control to evaluate the ethicality of recruitment policies and procedures.	<a href="#">Landrum and Chastain (1995)</a> describe a short “spot-check” form developed to assess the educational value of undergraduate participation in research. <a href="#">Leak (1981)</a> describes a 10-item post-study questionnaire for gathering information on students’ perceptions of both coercion and educational value in relation to their research participation.
5. Implement a Web-based experiment management system.	Automated experiment management systems: (1) save time and money; (2) reduce scheduling and record-keeping errors; (3) are convenient for student, instructors, and researchers to use; and (4) enhance experimental/methodological integrity.
6. Use uniform recruitment practices as much as possible.	Recruitment practices are a potential source of measurement error. Approached arbitrarily, recruitment practices may increase participant pool heterogeneity resulting in increased variability in measurements and consequently an increase in the likelihood of Type II statistical error. They may also complicate random assignment (for ways to address individual differences that impact the timing of research participation, consult <a href="#">Aviv et al., 2002</a> ). Participant recruitment practices are thus an important, although often overlooked, part of the development of theoretical knowledge.

and provide examples of designs for different types of social research. Here, we focus on the practical issues of recruiting participants for sociology experiments. These issues are similar to those in psychology but often involve unique challenges (discussed later). Before beginning a new experiment, researchers must make several decisions: (1) how to secure a pool of participants; (2) what restrictions might be needed on the type of people invited to participate; and (3) how to secure and maintain a pool of participants from which to draw. In addition, online experiments and online virtual worlds are starting to receive more attention in sociology, so we conclude this section with a brief overview of this emerging methodology and issues related to participation.

## A Volunteer or Required Participation

In practical terms, the central issue of an incentive to participate concerns participant motivation. Monetary payment in exchange for voluntary participation in an experiment is an efficient way to ensure that participants focus on the experimental task. This is particularly true if money is incorporated *as part of the experimental manipulation*. For instance, in an experiment in which participants are told that they are working with another participant and that their task is to determine how many problems they can jointly solve, researchers can motivate participants to try their best to solve the problems by telling them that their pay in the study depends on how many problems they and their partner solve. In such an experiment, participants will likely pay close attention to the contributions of others (typically desirable in sociological experiments) and focus on problem solving. In contrast, participants who receive course credit for their appearance may pay little attention to the task at hand. Requiring experimental participation for course credit virtually guarantees participant availability, but it involves ethical problems (discussed previously) and may decrease the quality of participation. The important point here is that using different incentives in the same experiment or across replications may unwittingly affect participant behavior, thereby complicating the interpretation and comparison of study results. As such, the issue of participant motivation as it relates to recruitment incentives should not be approached in an arbitrary way.

## B Selection Criteria: Who Will Participate?

It is helpful to gather basic information about future participants before contacting them for a specific study, and nowadays prescreening is easily handled with Web-based experiment management systems such as Sona Systems. Some experiments require participants who possess certain characteristics. When these characteristics are necessary for the successful completion of an experiment, researchers target people with those characteristics for participation. For example, consider the hypothesis that educational attainment (as a status characteristic) affects influence. An experimental test might include only first-year

undergraduates who are matched with a (fictitious) more educated partner in a “lower-status” condition and a less educated partner in a “higher-status” condition.

## C Methods of Recruiting Participants

Participants may be recruited to participate in sociology experiments in a variety of ways. The appropriate method may be determined by available resources or the aims of the experiment. The two methods detailed here are in-person recruiting and semester-commitment recruiting.

The primary method of making contact with potential research participants in sociology has historically been in-person recruiting. This method provides participants with a brief synopsis of the sort of phenomena being examined in the laboratory, an introduction to one or a number of researchers and research assistants conducting experiments in the laboratory, and the opportunity to provide researchers with a way to make contact in the future to schedule a date and time to serve as a participant in an experiment, which may include use of a Web-based experiment management system.

In-person recruiting requires a coordinated and well-organized team effort to visit high-enrollment classes. The first step is to create a list of high-enrollment courses. With the advent of electronic course enrollment, this process became relatively simple. Most institutions make enrollment information available before classes begin. The next step is to establish contact with the instructors of the high-enrollment courses on the list and request permission to visit their classes. E-mailing each of the instructors has proven efficient. There are several important points to convey in a message requesting a class visit. While specifying that a class visit would be most effective during the first week or two of classes, it is important to allow instructors the freedom to select the date for a recruitment visit. It is important to also mention to instructors that no class time is required for the visit but that having 5–10 minutes at the beginning of class would be most effective. Finally, it is important to mention that students will be told that their choice to participate in an experiment is voluntary and not related to the course or their grade in the course.

After receiving approval to make a class visit, a team of research assistants is assembled to assist with the visit. Generally, classes with 100 students require two or three research assistants. Larger classes (300–700) may require as many as five assistants. Prior to making the visit, a handout is created. Handouts include information about how to use a Web-based experiment management system to browse and sign up for studies, or if such a system is not used, they provide spaces for a prospective participant to fill in his or her name, gender, year in school, phone number(s), best time to be contacted, and a description of any experiments the student may have participated in previously. Information on past participation can be used to avoid redundant participation in the same or similar studies.

With enough handouts for each student enrolled in the class, the team of research assistants arrives at the classroom location approximately 10–15 minutes prior to the start of class. As students enter the classroom, research assistants distribute the handouts. Two or three minutes prior to the beginning of class (or the minute class begins if the instructor permits), one assistant stands at the front of the class, requests students' attention, briefly introduces the recruitment team, and explains the nature of research participation. The content of the description of research participation includes information concerning: (1) the types of topics being investigated in the laboratory; (2) the typical duration of studies; (3) whether monetary or other incentives to participate are being offered; and (4) the assurance that students are not obligated to participate in an experiment. Furthermore, it is critical to emphasize that participation or nonparticipation in an experiment in no way affects students' grade in the class (unless the instructor offers course credit for participation). After completing the presentation, research assistants quickly collect the handouts (i.e., when they are used to gather contact information) and exit the classroom. To increase the likelihood that instructors will continue to work with experimenters in the future, research assistants must take care not to disrupt class time. Efficient administration of the recruitment process is of paramount importance to maintaining an enduring participant pool. Another concern is that extra credit can be problematic if only some instructors consent to offering an incentive to participate while others do not. In that case, participants in the same study from different classes have different incentives to participate, which may, as discussed previously, unwittingly affect participant behavior in ways that might confound study results. Having discrepancies in participation incentives across classrooms is also ethically questionable.

If a Web-based experiment management system is not used, a few important steps in maintaining a pool of potential participants and scheduling participants for studies occurs *after* classroom visits are conducted. First, all of the contact information obtained via in-class recruitment efforts should be entered into an electronic database. The ability to access contact information electronically allows researchers conducting different experiments who use the same participant pool to avoid contacting potential participants more than once and to keep track of those who have already participated in an experiment. Establishing a password to protect potential participants' contact information adds a layer of security to this sensitive information. After the contact information for each potential participant has been entered manually into a database, the sheets containing that information should be destroyed. Second, scheduling participants for specific sociology experiments requires a dedicated scheduler. The individual filling this role references the database of potential participants and contacts them via telephone or e-mail. The scheduler's goal is to fill all available experimental sessions with participants from the list. Prior to calling or e-mailing potential participants, the principal investigators for each experiment receive approval of a "script" used by the scheduler to make appointments. Approval of a script is

granted by the university's IRB. The scheduler then uses the approved script to contact potential participants. In our experience, contacting participants during the weekend before the week they would be scheduled to participate is most effective. Phoning in the late afternoon is also effective. After the scheduler fills all possible experimental sessions for the upcoming week, making reminder calls to participants the day before they are scheduled to participate is good practice. We find that this strategy decreases the rate of "no-shows" for experimental sessions. Paid schedulers can also be offered a bonus for each scheduled participant who successfully completes participation in the study. Another useful strategy, albeit a more costly one, involves "overbooking" experimental sessions and arranging payments for those who come but do not take part in the experiment.

A second method of recruiting a pool of research participants in sociology is semester-commitment recruiting. As its name suggests, researchers using this technique secure a commitment from participants to actively participate in experiments for an entire semester. This method is only valuable in situations in which participants in the pool may take part in multiple experiments or in the same experiment multiple times. As such, this technique is not recommended for experiments in which deception is used early in the term.

Making initial contacts with potential participants for semester-commitment recruiting can be achieved by way of e-mails or the in-class technique. The primary way in which this method differs from others is that researchers obtain a commitment from participants to participate in as many experiments (or sessions of the same experiment) as they are able during a semester (i.e., as is reasonable given their schedule). If participants are paid for their participation, the pool of semester-commitment participants can be thought of as employees hired to contribute to the research being conducted. If participants receive course credit for their participation, these participants may be thought of as students enrolled in a semester-long course whose grade depends on the frequency of their participation. One way to manage course credit in lieu of payment is to create a research practicum course in which participants may register. Although participants in this framework may not receive instruction in the traditional format (i.e., with lecture and discussion), their experiences and observations of the research process justify college credits in the same way internship credits often count toward obtaining an undergraduate degree.<sup>2</sup>

## D Online Experimentation and Virtual Worlds

Although the technological and organizational issues and "best practices" for carrying out controlled social science experiments on the Web have only recently started to receive due attention (Kohavi, Longbotham, Sommerfield, & Henne, 2009), programs such as TESS (Time-Sharing Experiments for the

---

2. Although, as we discussed in the previous section, the *required* nature of such participation raises ethical concerns.

Social Sciences) provide researchers with a unique opportunity to participate in this new and exciting area of experimentation in sociology and related disciplines. Funded by the National Science Foundation (SES-0818839), the TESS program reviews brief, 5-page proposals from social scientists interested in conducting Web-based experiments with participants who are representative of the U.S. general population. Studies detailed in successful proposals are carried out at no cost to the principal investigator by the private research company, GfK (formerly Knowledge Networks). Population-based experiments such as those that can be conducted through TESS are particularly useful in cases in which the researcher seeks to combine the internal validity of experimentation with the external validity of probability sampling ([Mutz, 2011](#)).

In addition, popular multiplayer online role-playing games (MMORPGs) and other kinds of online virtual worlds represent another emerging frontier in social science laboratory research ([Lovaglia & Willer, 2002](#)). For example, the online virtual world “Second Life” has had more than 35 million “Residents” (i.e., players) since it was launched in 2003 (<http://gridsurvey.com>), many of whom participate on any given day to create a place for themselves in a virtual society in which social meaning and structures are created, negotiated, and modified as users of varying power and status interact through avatars, become involved in groups, and participate in an internal economy by exchanging a variety of goods and services with one another. The principal advantage of such Internet laboratories is that complex social situations can be followed by investigators and outcomes can be observed at various points as Residents play out their roles in an ongoing manner during the course of months or even years. For example, [Bakshy, Karrer, and Adamic \(2009\)](#) obtained data directly from the makers of Second Life (without personally identifying information) and analyzed complete Resident data over a 130-day period. They used these data to identify and model the influence dynamics underlying the diffusion of content (buildings, fashion, etc.) through evolving social networks. One drawback of this kind of research is that control over the characteristics and circumstances of participants is virtually absent. However, the drawback of participant heterogeneity is compensated by the huge number of participants and the amount of demographic information that can be collected to establish statistical control.

University undergraduates are not ideal for studying some research questions, as dictated by theory and sometimes practicality. Web-based experiments and online virtual worlds with simulated communities are a promising alternative for studying questions not amenable to analysis with recruits from university student populations.

## **IV PARTICIPANTS IN POLITICAL SCIENCE**

Political science studies research questions for which undergraduates often make suboptimal research participants. Participation is reserved for adults and undergraduates who have little or no previous experience as research participants.

A variety of techniques are used in political science to recruit participants for experimental research that could be classified as laboratory based, whether those experiments occur in a university laboratory, in the field, are survey based, or online. Typically in political science, potential participants are told they will receive some form of reward, usually monetary pay, for participation.

## A Laboratory Locations

Recruitment varies greatly even in laboratory experiments. As is often the case now in psychology and sociology, one method is to use a Web site where individuals can register and sign up for participation. One example is the Interdisciplinary Experimental Laboratory (IEL) at Indiana University (<http://www.indiana.edu/~ielab>), a joint endeavor of faculty and staff from political science, psychology, economics, and geography. Details on the variety of participant recruitment methods employed by the IEL are available on its Web site. Following the usual procedures, students at Indiana University first visit a Web site and register an account. They then sign up for an experiment by reviewing a dynamically updated calendar.

Another means of recruiting participants in political science involves visiting college dormitories to collect personal information from students who want to participate over the course of the year. This information can be entered into a database, and a selection of prospective participants may be generated from that list. The next step is to send an e-mail to each student directing him or her to a Web site to sign up for the experiment. [Wilson and Eckel \(2006\)](#) used this method to recruit participants to explore beauty and expectations in trust games.

Non-Internet methods to recruit participants have been used as well. In an interdisciplinary project (sociology and political science), [Sell and Wilson \(1999\)](#) recruited participants from introductory social science and humanities classes. Students were told they would be paid in cash for volunteering in “decision-making” experiments. Those who volunteered were scheduled at their convenience and randomly assigned to experimental conditions.

Another example of non-Internet recruiting is seen in the work of [Bottom, Eavey, Miller, and Victor \(2000\)](#). They recruited 240 participants from undergraduate and graduate classes in the school of business, the school of engineering, and the college of arts and sciences. They advertised an experiment in “collective decision-making” in classrooms, through an electronic bulletin board, and through sign-up sheets posted in the student union. All methods mentioned a minimum payment of \$3 plus an opportunity to earn more based on group decisions.

## B Laboratory Locations Using Nonstudent Participants

In laboratory experiments, when the student population is not desired, researchers may also draw from the general public. For example, [Berinsky and Kinder \(2006\)](#)

enlisted participants through posting advertisements and also recruited from local businesses and voluntary organizations. Participants reflected great diversity (compared to the college student sample), although as discussed previously, for theory-testing purposes this is not desired. In addition, [Ansolabehere, Iyengar, Simon, and Valentino \(1994\)](#) examined the effects of negative campaign advertising on voter turnout. During an ongoing political campaign (therefore featuring actual candidates and voters), they recruited participants by placing advertisements in local papers, handing out flyers in shopping malls and other public venues, posting announcements in employer newsletters, and telephoning people from voter registration lists. All participants were promised payment of \$15 for an hour-long study. Although the sample was not random, descriptive statistics suggested that it reflected the population from which it was drawn. Another study by [Iyengar, Peters, and Kinder \(1982\)](#) recruited participants from a specific city using classified advertisements that offered \$20 to those who participated in “research on television.” Interested citizens responded by phone and were randomly assigned to experimental conditions and scheduled at their convenience. Descriptive statistics suggest this method also produced a roughly representative sample of the city population. [Redlawsk \(2002\)](#) recruited participants in a large city by contacting different organizations (including the YMCA and a senior citizen center) and requesting that they invite their members to volunteer in experiments in return for a \$20 donation to the organization per member who participated.

## C Laboratory Experiments in the Field

In some field experiments, a community becomes the laboratory. For example, [Eldersveld's \(1956\)](#) often cited early work examined the effects of personalized versus impersonalized propaganda techniques on voting behavior. Eldersveld mailed out different forms of propaganda and followed up with post-experiment interviews. Local participants came out of a sampling frame of city clerk records. He selected all people living in four precincts of a central area and who had voted regularly in both state and national elections (but not in local elections, for reasons related to his research question). Although not perfectly representative, the sample size of 187 in two conditions allowed much statistical power for the use of statistical control variables.

[Gerber and Green \(2000\)](#) randomly selected households and exposed them to direct mailings, telephone calls, or personal appeals before a general election to determine which had the most impact on voter turnout. From a complete list of registered voters, they created a sampling frame of households. This technique generated a sample of 22,077 households. The effectiveness of randomization was checked using voter turnout data from an earlier election—a technique based on statistics and that showed there would be no significant difference between current and past voting behavior. The benefit of this technique is the large sample size that allows statistical control to overcome the loss of experimental control occurring with a heterogeneous sample.

Bahry and Wilson (2006) recruited participants for their field experiment using a sampling frame of individuals who had participated in an earlier interview pool in Russia. A total of 646 participants were included, with 252 from Tatarstan and 394 from Sakha. Experiments were conducted in small villages, medium-sized cities, and large urban areas within these Russian republics. Experiments were limited to villages and medium-sized cities where at least 20 individuals had been interviewed previously. Some medium-sized cities were skipped where travel was difficult or impossible. Payment for approximately 2 hours of participation reflected a week's wage or more for 62% of their participants.

Finally, Wantchekon (2003) conducted an experiment in the Republic of Benin in West Africa. Working with a team of consultants who helped him contact the leadership of selected parties, he communicated directly with them and campaign managers who then agreed to run an “experimental political campaign” in select districts. From his list of 84 districts, Wantchekon chose 8 districts and divided each into three subgroups. Each subgroup was exposed to either one of two experimental conditions or served as a control.

## D Survey and Online Experiments

In survey-based experiments, investigators use secondary data while adding a manipulation. Gilens (1996) did this to examine whether white Americans' opposition to welfare is rooted in prejudice against African Americans or nonprejudice reasons. Using the National Race and Politics Study data set—a national telephone survey—he applied a manipulation in the survey in which half of the respondents were asked a specific attitudinal question about whites, and the other half were asked the same question about African Americans. Nelson and Kinder (1996) also used a secondary data source to recruit participants and create an experiment. In their work, participants were recruited from the sampling frame of respondents who completed the 1989 National Election Study (NES) and who also had provided their telephone numbers. Randomly drawn from this frame, the researchers created a representative sample of the American adult population. Advantages of survey experiments are large sample size, the ability to randomly assign the respondents to questions, and the ability to generalize results to a larger population if desired. They also have disadvantages. Gaines and Kuklinski (2007) review the typical uses of survey experiments in political science and identify problems and solutions specific to this methodology.

Online experiments in political science are also performed using a variety of recruiting techniques. OxLab at the Oxford Internet Institute (<http://www.governmentontheweb.org>) maintains a database of research participants including both University of Oxford students and nonstudents from the city of Oxford. Margetts, John, Escher, and Reissfelder (2011) studied how information on the Internet affects political participation by recruiting 668 individuals from the OxLab database and having them participate remotely in a Web-based

experiment using their own Internet connection. Using a less active approach to participant recruitment, investigators may also rely on “drop-ins,” in which participants come across the experiment while surfing the Internet. Another method uses banner ads that offer some kind of incentive for participation. Finally, Iyengar (2002) has used a market research firm, Knowledge Networks, to reach a nationwide representative sample. Through standard telephone methods, Knowledge Networks recruits a continuous sample of individuals between the ages of 16 and 85 years who are provided free access to WebTV. In exchange, these individuals agree to participate on rotation in different studies. Iyengar examined online self-selection and found that drop-in Internet experiment participants reflect reasonably well the online user population, but participants still differ from the general population because non-Internet users are not reflected in the experiment sample. Iyengar also noted that among participants in online experiments, Republicans outnumbered Democrats and Independents compared to the broader online population. This is an important issue for political scientists and others who may prefer a “party-representative” sample for their research. In general, using the Internet as a platform for experiments offers many advantages (e.g., a worldwide geographic domain, the ability to reach diverse populations, and low cost). As with any format, however, there are drawbacks as well (e.g., sample selection bias, excluding the population with no Internet access, and lack of participant homogeneity for theory testing).

## V CONCLUSION

In describing the methods used by laboratory researchers in several social science disciplines to recruit and work with human participants, we hope to have gone into enough detail to allow interested researchers to begin research with human participants in their own laboratories. As we have noted, recent technological advances require an expanded definition of laboratory experiments to include theory-driven fundamental research carried out in a variety of physical settings and using a variety of participant interface and data collection techniques.

## REFERENCES

- Ansolabehere, S., Iyengar, S., Simon, A., & Valentino, N. (1994). Does attack advertising demobilize the electorate? *American Political Science Review*, 88(4), 829–838.
- Aviv, A. L., Zelenski, J. M., Rallo, L., & Larsen, R. J. (2002). Who comes when: Personality differences in early and later participation in a university subject pool. *Personality and Individual Differences*, 33, 487–496.
- Bahry, D. L., & Wilson, R. K. (2006). Confusion or fairness in the field? Rejections in the ultimatum game under the strategy method. *Journal of Economic Behavior and Organization*, 60(1), 37–54.
- Bakshy, E., Karrer, B., & Adamic, L. A. (2009). Social influence and the diffusion of user-created content. In *Proceedings of the 10th ACM conference on electronic commerce*, Stanford, CA. New York: ACM.

- Berger, J., M. Fisek, M. H., Norman, R. Z., & Zelditch, M., Jr. (1977). *Status characteristics and social interaction: An expectation states approach*. New York: Elsevier.
- Berinsky, A. J., & Kinder, D. R. (2006). Making sense of issues through media frames: Understanding the Kosovo crisis. *Journal of Politics*, 68(3), 640–656.
- Bernard, L. C. (2000). Variations in subject pool as a function of earlier or later participation. *Psychological Reports*, 86, 659–668.
- Bottom, W. P., Eavey, C. L., Miller, G. J., & Victor, J. N. (2000). The institutional effect on majority rule instability: Bicameralism in spatial policy decisions. *American Journal of Political Science*, 44(3), 523–540.
- Bowman, L. L., & Waite, B. M. (2003). Volunteering in research: Student satisfaction and educational benefits. *Teaching of Psychology*, 30, 102–106.
- Britton, B. K. (1979). Ethical and educational aspects of participating as a subject in psychology experiments. *Teaching of Psychology*, 6, 195–198.
- Calder, B. J., Phillips, L. W., & Tybout, A. M. (1981). Designing research for application. *Journal of Consumer Research*, 8, 197–207.
- Coulter, X. (1986). Academic value of research participation by undergraduates. *American Psychologist*, 41, 317.
- Daniel, R. S. (1987). Academic value of research participation by undergraduates: Comment on Coulter. *American Psychologist*, 42, 268.
- Davis, D. D., & Holt, C. A. (1993). *Experimental economics*. Princeton, NJ: Princeton University Press.
- DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, 48, 147–160.
- Eldersveld, S. J. (1956). Experimental propaganda techniques and voting behavior. *American Political Science Review*, 50(1), 154–165.
- Flagel, D. C., Best, L. A., & Hunter, A. C. (2007). Perceptions of stress among students participating in psychology research. *Journal of Empirical Research on Human Research Ethics*, 2, 61–67.
- Gaines, B. J., & Kuklinski, J. H. (2007). The logic of the survey experiment reexamined. *Political Analysis*, 15, 1–20.
- Gerber, A. S., & Green, D. P. (2000). The effects of canvassing, telephone calls, and direct mail on voter turnout: A field experiment. *American Political Science Review*, 94(3), 653–663.
- Gilens, M. (1996). “Race coding” and white opposition to welfare. *American Political Science Review*, 90(3), 593–604.
- Gil-Gómez de Liano, B., León, O. G., & Pascual-Ezama, D. (2012). Research participation improves students’ exam performance. *Spanish Journal of Psychology*, 15, 544–550.
- Hertwig, R., & Ortmann, A. (2008). Deception in social psychological experiments: Two misconceptions and a research agenda. *Social Psychology Quarterly*, 71, 222–223.
- Iyengar, S. (2002). Experimental designs for political communication research: From shopping malls to the Internet. *Work presented at the Workshop in Mass Media Economics, Department of Political Science, London School of Economics*, June 25–26.
- Iyengar, S., Peters, M. D., & Kinder, D. R. (1982). Experimental demonstrations of the “not-so-minimal” consequences of television news programs. *American Political Science Review*, 76(4), 848–858.
- Jackson, M. J., & Cox, D. R. (2013). The principles of experimental design and their application in sociology. *Annual Review of Sociology*, 39, 27–49.
- Jung, J. (1969). Current practices and problems in the use of college students for psychological research. *The Canadian Psychologist*, 10, 280–290.

- Kagel, J. H. & Roth, A. E. (Eds.). (1995). *The handbook of experimental economics*. Princeton, NJ: Princeton University Press.
- Kohavi, R., Longbotham, R., Sommerfield, D., & Henne, R. M. (2009). Controlled experiments on the Web: Survey and practical guide. *Data Mining and Knowledge Discovery*, 18, 140–181.
- Landrum, R. E., & Chastain, G. (1995). Experiment spot-checks: A method for assessing the educational value of undergraduate participation in research. *IRB: A Review of Human Subjects Research*, 17, 4–6.
- Landrum, R. E., & Chastain, G. (1999). Subject pool policies in undergraduate-only departments: Results from a nationwide survey. In G. Chastain & R. E. Landrum (Eds.), *Protecting human subjects: Departmental subject pools and institutional review boards* (pp. 25–42). Washington, DC: American Psychological Association.
- Leak, G. K. (1981). Student perception of coercion and value from participation in psychological research. *Teaching of Psychology*, 8, 147–149.
- Lovaglia, M. J. (2003). From summer camps to glass ceilings: The power of experiments. *Contexts*, 2(4), 42–49.
- Lovaglia, M. J., & Willer, R. (2002). Theory, simulation, and research: The new synthesis. In J. Szmata & K. Wysienska (Eds.), *The growth of social knowledge* (pp. 247–263). Westport, CT: Praeger.
- Lucas, J. W. (2003). Theory-testing, generalization, and the problem of external validity. *Sociological Theory*, 21, 236–253.
- Margetts, H., John, P., Escher, T., & Reissfelder, S. (2011). Social information and political participation on the Internet: An experiment. *European Political Science Review*, 3, 321–344.
- McCord, D. M. (1991). Ethics-sensitive management of the university subject pool. *American Psychologist*, 46, 151.
- Miller, A. (1981). A survey of introductory psychology subject pool practices among leading universities. *Teaching of Psychology*, 8, 211–213.
- Mutz, D. C. (2011). *Population-based survey experiments*. Princeton, NJ: Princeton University Press.
- Nelson, T. E., & Kinder, D. R. (1996). Issue frames and group-centrism in American public opinion. *Journal of Politics*, 58(4), 1055–1078.
- Nimmer, J. G., & Handelsman, M. M. (1992). Effects of subject pool policy on student attitudes toward psychology and psychological research. *Teaching of Psychology*, 19, 141–144.
- Padilla-Walker, L. M., Zamboanga, B. L., Thompson, R. A., & Schmersal, L. A. (2005). Extra credit as incentive for voluntary research participation. *Teaching of Psychology*, 32, 150–154.
- Redlawsk, D. P. (2002). Hot cognition or cool consideration? Testing the effects of motivated reasoning on political decision making. *Journal of Politics*, 64(4), 1021–1044.
- Roberts, L. D., & Allen, P. J. (2013). A brief measure of student perceptions of the educational value of research participation. *Australian Journal of Psychology*, 65, 22–29.
- Rosell, M. C., Beck, D. M., Luther, K. E., Goedert, K. M., Shore, W. J., & Anderson, D. D. (2005). The pedagogical value of experimental participation paired with course content. *Teaching of Psychology*, 32, 95–99.
- Rosenthal, R., & Rosnow, R. L. (1969). The volunteer subject. In R. Rosenthal & R. L. Rosnow (Eds.), *Artifact in behavioral research* (pp. 61–112). New York: Academic Press.
- Sell, J., & Wilson, R. K. (1999). The maintenance of cooperation: Expectations of future interaction and the trigger of group punishment. *Social Forces*, 77(4), 1551–1570.
- Sieber, J. E., & Saks, M. J. (1989). A census of subject pool characteristics and policies. *American Psychologist*, 44, 1053–1061.

- Sona Systems. (2013). *Product: Web-based subject pool management*. Estonia: Sona Systems. Retrieved August 21, 2013, <http://www.sona-systems.com/product.asp>.
- Stevens, C. D., & Ash, R. A. (2001). The conscientiousness of students in subject pools: Implications for “laboratory” research. *Journal of Research in Personality*, 35, 91–97.
- Szmatka, J., & Lovaglia, M. J. (1996). The significance of method. *Sociological Perspectives*, 39, 393–415.
- Thye, S. R. (2000). Reliability in experimental sociology. *Social Forces*, 74, 1277–1309.
- Trafimow, D., Madson, L., & Gwizdowski, I. (2006). Introductory psychology students’ perceptions of alternatives to research participation. *Teaching of Psychology*, 33, 247–249.
- Turner, J. H. (2006). Explaining the social world: Historicism versus positivism. *Sociological Quarterly*, 47, 451–463.
- Wantchekon, L. (2003). Clientelism and voting behavior: Evidence from a field experiment in Benin. *World Politics*, 55, 399–422.
- Wilson, R., & Eckel, C. C. (2006). Judging a book by its cover: Beauty and expectations in the trust game. *Political Research Quarterly*, 59(2), 189–202.
- Witt, E. A., Donnellan, M. B., & Orlando, M. J. (2011). Timing and selection effects within a psychology subject pool: Personality and sex matter. *Personality and Individual Differences*, 50, 355–359.
- Wright, D. B. (1998). People, materials, and situations. In J. Nunn (Ed.), *Laboratory psychology: A beginner’s guide* (pp. 97–116). Hove, UK: Psychology Press.

## Chapter 6

# Developing Your Experiment

Lisa Slattery Walker

*University of North Carolina—Charlotte, Charlotte, North Carolina*

## I INTRODUCTION

For decades, at least, social scientists have discussed *whether* we should do experiments, *why* we do experiments, and even *when* we do experiments. But rarely do we discuss—particularly in print—*how* we do experiments. Reports of experimental research do not regularly describe the myriad minor and major decisions that went into the project. However, a social science experiment is made up of many details, and most of the details have to be addressed competently or the outcomes of the experiment can be useless for the purposes intended or, worse, misleading. With this chapter, I hope to remove some of the mystery from conducting social scientific experiments by presenting a few particulars about how one should design, conduct, and analyze results from an experiment in order to maximize the usefulness of its outcomes.

Elsewhere in this book, you can read about certain elements of how to conduct an experiment (e.g., technological issues, ethical concerns, training those who will be conducting your experiment, recruiting participants, and maintenance of records), but I focus on the initial design of an experiment and on the development of that design. Especially for new investigators, translating abstract considerations of good design into an actual, workable experiment can seem a daunting task. I hope this chapter can help with the process of doing a real experiment. I hope also to make clear some elements of experimental design that are often overlooked in more philosophical or abstract discussions.

For convenience, I present three steps in creating and conducting an experiment: (1) *designing the experiment*; (2) *pretesting* the operations and *pilot testing* the experiment; and (3) *analyzing and interpreting the data* it produces. Each stage in the execution presents challenges and requires decisions on the part of the experimenters. As an overview, it is helpful to keep in mind that the essence of experimental design is to create a situation (or possibly multiple situations) that includes *all* the factors described in a theory, and *only* those factors, in order to test ideas from the theory. In most cases, an experiment will contrast

multiple conditions, and those will ideally be identical to each other except for differences required by contrasting hypotheses.

## II DESIGNING THE EXPERIMENT

Good experiments begin with an explicit theory, which has the structure to permit predictions of derived consequences. Theoretically derived consequences are sentences (hypotheses) detailing outcomes that a theory predicts, given a specified kind of situation. However, derived consequences usually contain abstract theoretical terms, not concrete terms that are immediately observable.

For instance, a derived consequence of David Willer's network exchange theory (NET) (e.g., Willer et al., 2002) might be the following: "A person occupying a central node in a network will have more negotiating power than someone occupying an isolated node." Although the sentence's meaning may be clear, it does not tell us in terms of operations just what being "central" means or how to observe "power." On the other hand, "A person with two potential exchange partners will gain more points in negotiation than someone with only one partner" translates the theoretical terms into observable facts in an experimental situation. The first sentence is a derived consequence of a theory; the second is a testable hypothesis. In designing an experimental test of NET, it is necessary to create such testable hypotheses.

No experiment can test all of the derivations of a theory; one must choose some of those derived consequences for hypotheses, preferably a set with some range of theoretical assumptions. For instance, if the theoretical foundation of the experiment is a theory having five general propositions, it is wise to examine which propositions are used in the derivation of each derived consequence. Usually, any two, three, or four propositions will yield many derivations. The experimenter must choose to test a few derivations from among a large set. Although the choice is somewhat dependent on personal preference and empirical simplicity, it is usually wise to be sure that the experiment tests as many of the propositions as possible. Thus, testing two derivations that both are implied by propositions 2 and 3 is redundant, and if no derivation that uses propositions 4 and 5 is tested, the experiment will provide only a partial test of the theory.

The design task is then to translate the conceptual terms in which the theory is couched into a realistic, although not usually real-world, situation in which the experimenter controls many of the elements. By "realistic," I mean that the situation must be understandable to the participants, and it cannot be so bizarre that they feel they have entered the "Twilight Zone." On the other hand, an experiment is not a natural setting. If a natural setting existed that provided a good test of theoretical derivations, there would be no need to design the experiment. Thus, most experiments seem a little strange to participants, but as long as they understand the important aspects of it and their behavioral options, realism is neither needed nor desirable. The more "realistic" an experiment seems, the more likely that some (but probably not all) of the participants will fall into

familiar role behaviors in it. If an experiment reminds some participants of a high school classroom, for instance, they may activate ways they typically behave in classes: some will be attentive, some defiant, some bored, etc. Those role behaviors and the variability across participants, of course, are not what an experimenter wants. What she wants is for the situation to present all and only the previously set initial conditions of her design. The situation should seem real to the participants in that it is understandable and it has consequence, although it may be unlike anything they have ever encountered before.

The important practical consideration is how to make the experimental situation understandable and relevant to the participants, with thought to their culture and background, while staying true to the theory. A number of abstract design elements thus come into play, particularly variables and conditions, manipulations, and manipulation checks.

## A Standard Protocols

Many theoretical programs develop standardized experimental protocols. If you happen to be working in one of these areas, such a protocol is an excellent tool for developing your own experiment. In addition to providing you with what amounts to a shortcut to a good experiment, the use of such standardized protocols increases the likelihood that your results will be meaningful in an ongoing scientific conversation.

Changes made to standardized protocols should be driven by theoretical questions. That is, you should only alter the elements of the design needed to test your hypotheses. Making other changes to the protocol for reasons such as “it seems better” or “it will be easier this way” often leads to unintended, and generally unmeasured, consequences.

When you do find it necessary to change an established design, be sure to carefully pretest any altered elements. Later, I discuss in detail the importance of pretesting, but in the case of standard protocols it is particularly important. When you use a standard protocol, you increase the comparability of your results to others that use the same protocol. When you alter the design, you may just decrease that comparability. However, others will certainly make comparisons anyway, and it is your responsibility to ensure that they are valid.

## B Variables and Conditions

I discuss, in turn, a number of abstract design considerations with which researchers must deal when conducting a social scientific experiment. Primary among these considerations are manipulations, where the researcher puts into motion the initial conditions and independent variables as specified by his or her theory. It is important that the researcher is clear from the outset just what are the independent and dependent variables in the hypotheses, and which ideas are being tested in the experiment, in order to create the experimental conditions.

In other words, it is important to be clear just which ideas from the theory are to be tested and what sorts of situations are appropriate for that purpose. What sorts of situations the theory describes are often couched as scope conditions or, sometimes, limiting conditions.

Hypotheses (like derived consequences) can usually be stated in the form “If X, then Y.” More completely, they state, “Given a situation of a specified sort, if X then Y.” The first part of that sentence, a situation of a specified sort, describes the *initial conditions* of the situation. This governs the kind of experimental situation the researcher will create. (Of course, the experimental situation must also instantiate the scope conditions of the theory, as discussed by Foschi in Chapter 11.) The “X” represents the *independent variable(s)*. This is the element that will be introduced in some experimental conditions and not in others, or introduced at different levels in different conditions. The “Y” is the *dependent variable(s)*. This is what the researcher will measure once he or she has devised a suitable measurement instrument.

Once the variables are clear, one can determine how many experimental conditions are required. This requires a good understanding of the variables and the relevant number of levels each has. An incomplete number of conditions can be the downfall of an otherwise well-conducted experiment if the important comparisons cannot be made with the data. This includes the problem of not having baseline conditions when needed. However, not every experiment requires a full crossing of all of the levels of the independent variables in order to have a set of conditions that is complete for the purposes of testing the theoretical hypotheses under consideration. Again, it is important to be clear about exactly what one is testing when designing the experiment.

Knowing the predicted relationships among the values on the dependent variables is also important. Again, the theoretical concerns allow one to determine just what the relevant comparisons are across conditions. The design of the experiment should allow for, and lead inexorably to, making comparisons among conditions that will give a true and meaningful test of the hypotheses and therefore of the theoretical concepts.

## C Manipulations

Manipulations are the process by which an experimenter creates the independent variables operationally within the experimental setting. Manipulations in social science experiments frequently fall into the category of information that is given to the participants about themselves, anyone with whom they might be interacting, the situation, the task, or the social world. Other types of manipulations include the behavior of others in the situation (often computer programmed or performed by a confederate) or an imposed social network or structure.

The process of manipulations often includes a cover story or process of setting the stage as the researcher creates the scope and initial conditions and independent variables in an experimental condition. This cover story may or

may not include deception. Often, it is through this cover story that the experimental manipulations are made. It often comes in the guise of the instructions that participants receive regarding their participation in the study—what they are to do, when, how, and with whom. Manipulations can come in the form of commission—what is said in the given condition—or omission—what is not said in one particular condition that is said in others.

It is best to make sure that participants hear all of the relevant pieces of information at least three times during the cover story. As a rule, experimental participants are not especially attentive, and they often miss crucial pieces of information if it is only said once or even twice, so three times is required. They are also not usually very suspicious. Thus, although they might find the repetition of hearing something three times slightly tedious (if they notice it at all), it rarely causes them to disbelieve the cover story. It is better to err on the side of saying things too often, even with a risk of irritating the participants, than to err on the side of not saying things often enough and failing to properly create the conditions needed to create useful data.

(Please notice that I wrote “three times” three times in the preceding paragraph. If you even noticed the repetition, did it bother you? Probably not—and you are attending to the topic right now. Experimental participants certainly are not bothered by this kind of repetition, especially when it does not take place in just one paragraph!)

Here is an important rule about creating experimental manipulations: **SUBTLETY IS OUT OF PLACE IN EXPERIMENTAL DESIGN.** I trust I do not need to repeat that point. Sometimes investigators try to create subtle manipulations in a misguided attempt to preserve “naturalness” or, they think, to avoid drawing participants’ attention to hypotheses under test. The problem with subtlety is that it goes against the goal of creating a situation that instantiates conditions and variables of the hypotheses. Subtle elements of a situation are missed by some people and can be interpreted differently by different people. That means that if the manipulations are subtle, some participants will fail to notice them (and thus will not be in the situation the researcher thinks they are in), and some will interpret them differently. Both those effects will introduce variance in the data because people will be responding to different sorts of situations.

For instance, I once heard about an experimental design in which the researcher was interested in whether white participants would play a competitive game differently when they thought their opponent was white than when they thought the opponent was black. Because participants would never see the opponent, who in fact existed only as a computer program, the researchers intended to identify the opponent’s skin color by giving him what they thought was a “typically black” or a “typically white” name. These researchers wrote that they did not want to explicitly identify the opponent’s ethnicity for fear that it would activate either stereotypes or concerns about appearing egalitarian.

The problem is that the researchers do not know whether participants in the experiment will code the names they chose as revealing ethnicity; probably

some would and others would not. Worse, many participants may not even attend to the name of the opponent; who cares about his name if we are never going to meet? If it is important that participants classify their opponents in the game, then good experimental design makes that element unmistakable.

Generally, the more full a picture the researcher can paint of an element, the better the design. For instance, in the preceding design, the researchers could identify partner's skin color with an instruction such as "Your partner today is named \_\_\_\_\_. He is, like you, a white student here at State University." That, at least, is clear and unambiguous. However, to really activate any behavioral tendencies participants may have so that they may be seen in this situation, it would be even better to show a photograph, or a videotape with action and speaking cues, to instantiate this variable. The clearer and the more complete the instantiation of important design elements, the better.

Knowing who the participants are, in terms of their background and culture, is also useful in creating the manipulations of an experiment. Making the situation presented in the cover story relevant to the participants creates a more believable situation and one that they are more likely to take seriously. Students at elite universities, for example, might be more motivated by studies presented as furthering basic science, whereas those at less elite schools may be more focused when the study purports to help them learn something about themselves. Participants who are not used to the laboratory setting may need more friendly and repetitious instructions.

## D Manipulation Checks

One of the greatest strengths of laboratory experiments is the control the researcher has over the independent variables. However, researchers often fail to fully realize the potential of their experiments because they do not create the situation they intended to create. Careful experimental manipulation is important but not difficult. One tool all researchers should employ is the manipulation check.

Manipulation checks can take several forms. During pretesting (discussed later), experimenters should be sure to discuss with participants what they heard, how they interpreted it, and how it affected their behavior. In addition, a part of all experiments should be a questionnaire or interview (or both) in which the participants are asked about the experiment. The experimenter should verify that the information given to participants during the cover story was heard correctly and believed. This check should include any embedded information about partners, the task, the situation, or any other manipulations.

## III “THE GENDER EXPERIMENT”: A PRACTICAL EXAMPLE OF ABSTRACT CONSIDERATIONS

It may be easiest to understand how these abstract considerations look in an experiment by examining an actual experiment that has been conducted. I discuss

an experiment I designed and conducted (Rashotte, 2006). I use an example of my own work not out of ego but, rather, because it is only for an experiment of my own that I will know fully the considerations and decisions that went into its design, pretesting, and conduct.

I designed an experiment to examine how to control status beliefs associated with gender. This experiment was intended to test several related hypotheses from the status characteristics branch of the expectations states theoretical research program within sociology. I was not concerned with *whether* gender was associated with status for my participants; rather, I wished to demonstrate that *when* status beliefs were present, they could be controlled through certain mechanisms described in the theory. Thus, I made sure that status beliefs—favoring men or favoring women—were present in every condition of the experiment.

The theory posits a number of mechanisms that might allow general status beliefs to be overcome. I tested two in this study: (1) by presenting status information about a task that contradicts the generally held status beliefs (e.g., saying that women are generally better at the task at hand); and (2) by providing specific evidence that the generally held status beliefs do not hold for these individuals (e.g., saying that although men are generally better at the task, in this case the particular male is not very good at it and the woman is exceptionally good at it).

## A Standard Protocols

To design this experiment, I began with a design that has been used in dozens of previous experiments and thus had a number of known properties, a variant of the standard experimental situation described by Berger in Chapter 12. The task at hand, the delivery of experimental instructions, and the cover story have been well-established over decades of research. Technological advances have allowed for recent improvements as well.

## B Variables and Conditions

My independent variables were gender of participant, status information regarding gender and performance on the task, and performance feedback. Participants always interacted with purported partners of the opposite gender. In some conditions, participants were told that males would do better at the task to be completed; in others, they were told women would do better. In certain conditions, participants were given (fictional) feedback on a pretest.<sup>1</sup> I was interested in comparing the effect on my dependent variable of the general information that

---

1. All participants completed the short trial version of the task in order to maximize comparability among the conditions. Only in the “feedback present” conditions were participants given (fictional) scores for the trial version.

women did better versus the specific feedback that the female partner did better (but not how the two combined). I thus needed six conditions:

- Male participants, told males generally do better, with no feedback
- Male participants, told females generally do better, with no feedback
- Male participants, told males generally do better, but with feedback that the female partner did better
- Female participants, told males generally do better, with no feedback
- Female participants, told females generally do better, with no feedback
- Female participants, told males generally do better, but with feedback that the female participant did better than her male partner.

My dependent variable was how often the participants deferred to their partners in making decisions on the task when the partner disagreed with the participant.

## C Manipulations

Participants were brought into an isolated room containing a computer monitor, a television, and a video camera. They were told that the study would begin when everyone was settled into the various rooms (leading them to believe there are real other participants nearby)<sup>2</sup> and that, when the time came, they would need to look into the camera to introduce themselves.

The participants then saw a videotape of instructions (said to be live via closed circuit television). The instructions were presented by a “Dr. Gordon” who claimed to be an expert in the task at hand and the ability underlying good performance at that task. The tape included a “live” introduction from their “partner” and a chance for the participant to introduce him- or herself, at which time the participant appeared on the television in the room. The introductions included information about the school attended (always our institution, to equate on that status variable)<sup>3</sup> and hobbies, to make the partners seem more real to the participants.

The instructions delivered all three of the independent variables. The gender of the partner was first introduced when Dr. Gordon said, “I see we have two people working together today, a man and a woman,” and reinforced by seeing the partner on screen and by the partner reporting gender stereotypic hobbies. The status information (“Previous studies have shown that men/women are generally better at this task”) was repeated three different times, including once just before the data collection phase began. The feedback information was provided

---

2. In fact, I usually did conduct several participants at the same time in order to support the belief that they were interacting with a real other, even though in reality each was interacting with the same fictional partner.

3. Participants were also similar to their partners on race and age in order to eliminate other status effects.

in those two conditions by a colleague of Dr. Gordon's, "Ms. Mason," who was an expert at scoring performance at this task. Ms. Mason repeated the scores, and their meanings (unusually high or unusually low), three times. Ms. Mason was also videotaped but said to be just down the hall.

## D Manipulation Checks

I conducted several kinds of manipulation checks. The interviews mentioned previously were conducted in order to verify that the participants heard all of the relevant information regarding the independent variables and that they understood the task they completed. The interviews were also used to determine if the scope conditions of the theory were in place. During a pilot testing phase of the study (more on this later), these interviews were even longer and covered other topics, such as whether "Dr. Gordon" was pleasant yet scholarly, if the session was an appropriate length, and how much of the instructional detail the participant could recall (beyond the basics related to the independent variables).

Participants also completed a questionnaire just prior to the interview. The questionnaire served as a double check to the interview and also provided some guidance for the experimenter in terms of where there might be some problems with a particular participant. The questionnaire covered factual information as well as impressions and affective responses. Extreme emotional responses can indicate a participant for whom the study was problematic and not properly prepared, and they can be competing processes to the ones of theoretical interest in the study.

## IV PRETESTING AND PILOT TESTING

In addition to the abstract considerations described previously, pretesting and pilot testing are important elements of good experimental design and conduct that are, unfortunately, sometimes overlooked by researchers. Pretesting involves examining certain elements of the experiment in isolation; pilot testing involves conducting complete experimental sessions with an eye to what is and is not working as expected.

Both pretests and pilot tests are different from actual experimental sessions in that the participants are required to act as informants. They let the researchers know what works and what does not work in the cover story, the task and/or interaction situation, the data collection, and all other parts of the experiment. This information can be gathered through questionnaires, interviews, or free-response surveys; ideally, it is obtained from several methods. They allow the researcher to fix unanticipated problems in the design.

The main elements of the design that need to be examined during both pretesting and pilot testing are scope conditions of the theory, the initial conditions of the experiment, and the instantiation of the independent variables. Researchers need to find out from the participants if the scope conditions are

holding in order for the theory's predictions to have any validity. The initial conditions, the cover story, must be understandable and believable for the data to have any value. Thus, for example, some groups of participants may require that information be repeated more than three times to be comprehended. The independent variable(s) must be clear and reasonable to have an effect on the dependent variable(s).

In addition, researchers must ascertain that the measures are working as expected. Participants must be paying attention, so the task must be somewhat interesting to them. Usually, experimenters wish to have a task that challenges participants without being so difficult as to cause undue frustration. If participants become distracted or emotional, their behavior may reflect those effects rather than effects of the theoretical factors as expected by the experimenter. The measures—especially key measures of the dependent variables—must be both valid and reliable.

Participants must buy into the cover story, the situation, and the task. They must also believe that the experiment has some importance—if not to them personally, then to the researchers and to science generally. Cultural factors come into play here. Experimenters must determine, through pretests and pilot tests, how to frame the situation in order to get their participants to believe it and want to take it seriously. Different populations of participants will require different frames for the cover story, and often it is not possible to know how participants perceive the encounter until pretests and pilot tests are conducted.

It is also in the process of pretesting and pilot testing that experimenter effects (discussed more fully later) can be identified. By having multiple experimenters conduct pretests and pilot tests, with thorough double checking, experimenter expectancy and observer effects can be identified early, before they can contaminate the data collected. Technology can be introduced when needed to reduce experimenter–participant interaction. Double-check systems can be implemented, and training can be increased if necessary. (For more on training, see Shelly, Chapter 4, this volume.) New measures, less participant to observer effects, can also be introduced if other steps are not effective at minimizing experimenter effects.

When pretests or pilot tests show things are not working as expected, experimenters must determine what to fix and how to fix it. It is much easier to determine *that* things are going wrong than it is to determine just *how* things are going wrong. Pretesting various elements of the design, prior to starting pilot tests or after, can allow experimenters to isolate where the problems are occurring. Thorough interviews with participants, with specific questions related to important elements of the experiment such as scope and initial conditions, can also help pinpoint the issues.

Once identified through pretests or pilot tests, problems must be fixed by the researcher. Issues with scope and initial conditions can usually be corrected by altering the cover story to resonate more with the participants. Issues of hearing, comprehending, or believing the independent variables can be corrected

by rewording—and/or reiterating—the information. If the measures of the dependent variable are not working as expected, new measures can be added to or used to replace existing measures. If participants are reacting in an emotional way that is distracting them from the experiment, additional information must be provided to the participants to help them contextualize what they are experiencing.

## A Pretesting

Pretesting most frequently is used for experimental instructions, but it can also be used for tasks, confederates, and instruments. These elements are isolated from the rest of the experimental setting, and participants are asked to evaluate them independently. The important considerations are often the means of conveying the situational definitions and all of the interaction elements. Pretesting is essential to ensure that the abstract and theoretical concerns are translated into a practical reality for the participants.

Again, I use an example from my research to illustrate details about pretesting. In a different but related experiment from the one described previously, new actors were used to portray “Dr. Gordon” and “Ms. Mason.” To be sure that this new portrayal created the right situation, and created the desire on the part of the participants to be focused and serious, pretests were conducted just on the videotape of instructions (Rashotte, Webster, & Whitmeyer, 2005). In fact, only a short 10-minute segment of the tape was tested (this is long enough to get a sense of the situation but not so long as to require students to be brought into the lab to view it).

Dr. Gordon and Ms. Mason were intended to be authoritative and pleasant and to hold the attention of the participants. The short segment of tape was shown to students in classes at the same university where the experiment was to be conducted. Students also saw a 10-minute segment from another experiment previously conducted in which the “host” experimenter on camera had been shown to be effective at creating the right situation. The previous segment was sort of a control condition. By that, I mean it had been used successfully in prior research, and our question was whether the new segment was at least as good as the established one.

Students rated each person they viewed on 40 seven-point semantic differential items. An ideal answer for each item was determined by the researchers (though not shared with the students doing the rating). A final open-ended question was also presented to allow the students to raise any issues that might not have already been covered. The 40 items were classified into four general categories: authority and competence; absence of distractions; clarity; and serious manner. Comparisons between the mean ratings for each individual to the ideal rating showed that the new Dr. Gordon did not perform as well as the previous experimental host, but the new Ms. Mason performed satisfactorily.

The particular failings of the new Dr. Gordon indicated that the problem was with the demeanor of the actor and not with the instruction script. Thus, a new

tape was produced with a different actor (in fact, it was the same actor who was in the tape from the previous experiment). This tape was also pretested, and the ratings were then compared with those from the previous experiment and the original actor. Things were then satisfactory, and the experiment could proceed. Since then, we have used the pretest technique on additional actors playing the experimenters' roles and have found good results; these new actors can be used for future experiments using this design.

## B Pilot Testing

Once the various elements of the experiment have been pretested, pilot tests can begin. Pilot tests are complete experimental sessions, designated “test groups” or “test sessions,” in which the researcher spends additional time questioning the participants about their participation. Pilot tests can identify any problems that did not arise in pretests because pretests often only examine segments of the experiment in different settings. Pilot tests are like dress rehearsals in the theater: everything is done together, in sequence, to judge how it looks as a whole. Also like rehearsals, sometimes results show that minor changes are needed, and other times everything is fine and the design can be used for the rest of the run of the show or of the experiment.

It is not until pilot tests that competing processes are usually discovered. Competing processes include fatigue, hostility, and withdrawal. Experiments that are too long—from the participants’ points of view—or take place at the wrong time of day can lead to fatigue, which can cause the participants to be less focused and less serious about the session. When part of the cover story involves providing the participants with information about themselves or others in terms of abilities or other such characteristics, emotional responses can occur. Sometimes this leads to anger and hostility; sometimes it leads to sadness or other negative affect and withdrawal on the part of the participants. If these emotional reactions are not a relevant part of the experiment (i.e., if they have nothing to do with the theoretical derivations under test), they can be distracting and lead to corrupted data. Thorough questioning of pilot test participants can lead researchers to detect competing processes.

After pilot tests are completed and all identified problems addressed, the actual run of the experiment can begin. Once problems are fixed, sessions may be called “experiment” rather than “test groups.” If no problems arise and no changes are made to the procedures of the experiment, the “test groups” can be reclassified “experiment groups” retroactively.

## V ANALYZING AND INTERPRETING DATA

The final stage of an experiment is the analysis and interpretation of the data it produces. Here, I do not address statistical methods for experimental data generally because those have been well covered elsewhere and most social scientists

are well versed in them. However, there are two elements of data interpretation that I believe are frequently overlooked by researchers. First, power analyses are often skipped altogether, which may lead to researchers missing evidence that their hypotheses are supported, even in an otherwise excellently designed experiment. Second, experimenter effects must be considered during data interpretation in order to rule out competing explanations for one's findings.

## A Power Analyses

Statistical power analyses are easy calculations that allow one to determine the number of participants that will be required in an experiment in order to reliably detect meaningful differences in the dependent variable. Calculators to determine statistical power are readily available online.

Statistical power is best thought of as the likelihood of not making a Type II error (failing to reject a null hypothesis that is not true). As you reduce the chance of making a Type II error, you increase the statistical power and thus the test is more sensitive (Keppel, 1991). The likelihood of a Type II error can be lessened by having a sufficient number of participants in the experiment and reducing the variability within conditions.

Statistical power depends on three factors: the significance level  $\alpha$  (representing the probability of making a Type I error, or rejecting a null hypothesis that is true); the magnitude of the differences across conditions on the dependent variable(s); and the sample size  $n$  (Keppel, 1991). Most often, researchers are only concerned with the sample size because the effect sizes are predicted by the theoretical constructs and  $\alpha$  is set by convention.<sup>4</sup> Thus, many of the statistical tools that have been developed, including those online, are geared toward determining needed sample sizes.

Researchers should conduct power analyses to ensure that the data will be useful. If one does not have enough participants in each condition in order to detect the differences between the conditions on the dependent variable, then all will have been for naught. The calculation of "how many is enough" requires knowing the expected differences between conditions, the variability within conditions, and the desired level of significance.

For example, let us think about a simple experiment. This experiment has only two conditions, and the dependent variable is measured as a proportion. The value of the dependent variable for each condition will be compared to a fixed value, 0.60. The predicted mean of the dependent variable is 0.65 for condition 1 and 0.54 for condition 2, with a standard deviation in each condition of

---

4. The significance level that is used varies across disciplines and also according to the kinds of data available. Most experiments in sociology and economics currently use 0.05; psychologists also use 0.05, and sometimes 0.01, as do researchers in education. In political science experiments, 0.05 is often used, although for nonexperimental work in which samples may be smaller, 0.10 or 0.15 is sometimes used.

0.15. We use the traditional 0.05 alpha level and a beta level (statistical power) of 0.50, which is also conventional.

By entering all of this information into an online calculator,<sup>5</sup> we find that we need 24 participants in condition 1 to detect the difference between 0.60 and 0.65 with a standard deviation of 0.15. We also find that we require a sample size of 17 in condition 2, where our predicted value is more different from our comparison value. The more different the values we are comparing, all else being equal, the smaller sample sizes required. If we have decided *a priori* to collect data from 20 participants in each condition, we would not be able to determine if our predictions about the dependent variable were correct.

Statistical tests are available to allow one to determine sample size for comparisons to a fixed value (as discussed previously), for comparing two groups to each other, for comparing groups in ANOVAs, for regressions, for surveys, and for many other situations. It is also possible to compute statistical power and confidence intervals for a given sample size. For those contemplating truly complicated experimental designs, power calculations are even available for Poisson distributions, Latin square designs, and survival analyses.

## B Experimenter Effects

Another concern in creating useful data is experimenter effects, meaning how the experimenter behaves and how participants respond to the experimenter. Experimenter effects have been found in studies on a wide range of topics, from religious attitudes (Hunsberger & Ennis, 1982) to sex (Winer, Makowski, Alpert, & Collins, 1988). Experimenter effects may include differences in the ways different experimenters in a team handle the experiment or attempts by participants either to please or to annoy the experimenter. Whenever experimenters are creating a situation through manipulations, it is necessary to anticipate, recognize, and deal with experimenter effects.

Most social scientific experiments are at least single-blind, in that the participants are not told the study's hypotheses and they do not know which condition they are in. They will know only the information they have been given—not how it systematically varies from the information participants in other conditions are receiving or even that it is systematically varied from information in other conditions.

However, very few social science experiments are double-blind; usually the experimenter knows the assigned condition for each participant. Often, this is unavoidable because the design of the experiment requires different actions on the part of the experimenter in various conditions. In this case, however, steps must be taken to avoid experimenter effects that might bias the results of the

---

5. Here, I use one available from HyperStat Online, found at <http://davidmlane.com/hyperstat/analysisf.html>. Many others are of course available, but they produce comparable results and I have found this one to be especially easy to use.

experiment. Experimenter effects can be one of two kinds: observer/interpreter effects and expectancy effects.

Observer or interpreter effects can occur when the dependent variable requires judgment of some kind on the part of the experimenter, but the experimenter is not directly interacting with the participant at the time the judgment is made.<sup>6</sup> For instance, experimenters might be interested in coding instances of anger or of interpersonal influence from videotapes of discussion groups. No matter how detailed the instructions for coding may be, and no matter how well trained the experimenters are, there is always some ambiguity in recognizing such concepts in an actual discussion. It is possible, in those cases, for the experimenter to see what he or she expects to see based on the hypotheses.

To reduce this effect, it is often helpful to have a double check on the data, done by another researcher blind to the condition. This double check can be done for all participants or, when that is not feasible, for a randomly selected set of participants. Then reliability of the two coders can be compared and, when a satisfactory level is attained, the researcher can be more confident that the data accurately record phenomena of interest. The double check is especially useful when conducted during pretesting, as discussed previously.

A variant kind of reliability can be calculated in experiments in which experimenters are called on to make decisions in fairly complex situations, such as deciding after an interview whether a particular participant met all the scope and initial conditions of the experimental design. Again, the presence of ambiguity and the requirement for judgment make it important to assess whether interviewers are applying criteria uniformly. One way to assess that is to determine whether, for instance, they are classifying approximately the same proportions of their interviewees in the same ways. Natural variation among participants will make the proportions vary somewhat, but, on average, classifications should be fairly uniform across interviewers. As always, the researcher will have to decide how much variability across interviewers is acceptable and what level suggests further inquiry, perhaps with additional training of interviewers.

Expectancy effects—those showing up in participants' interpretations and behavior—can be much more problematic, and often more subtle, than observer effects. They may occur because of participants' desire to please the experimenter, to behave in ways they think he or she wants them to do. On the other hand, participants might also try to act in ways they think the experimenter does not want. Neither is desirable.

Trying to please the experimenter is actually more common than trying to annoy her. Experimental participants may be paid volunteers, and they usually want to help “science” or at least the authority figures conducting the experiment. Although trying to confirm the experimenter's hypothesis might seem like a good way to please her, it is by no means an easy thing to do. As noted

---

6. This is the case when the experimenter is viewing the participants on camera or via computer while the experiment is in progress or when the data are recorded in some manner for later coding.

previously, participants are not told what the hypotheses are, nor how experimenters think or hope they will behave. However, another way to please the experimenter is available and quite common—that is, to present a positively evaluated self in the situation. A participant wanting to make a good impression may treat an interaction partner more cooperatively than he or she otherwise would or might take a long time filling out a questionnaire and so “overinterpret” the questions.

Trying to annoy the experimenter is the other side. If participants believe they have been mistreated—perhaps by being coerced into giving a certain number of hours of research as a class requirement—they may want to show their displeasure by taking things out on the experimenter. Here, interestingly, they may well try to disconfirm hypotheses of the study. Although they cannot be sure what the hypotheses are, they can be reasonably assured that if they act in a bizarre manner, it will be disconfirmatory.

Thus, trying to please the experimenter is likely to generate unusually prosocial behaviors; trying to annoy the experimenter is likely to generate very odd behavior. If an experimenter suspects either of these is a significant factor in the experiment—which, I hope, will be determined in pretesting—it will be important to take steps to deal with it. Generally, an experiment should not be generating negative emotions and hostile behaviors. If those appear, it is worth taking the time to interview participants at length to learn the sources of the feelings and then take appropriate steps to reduce or eliminate them. Paid volunteer participants are likely to feel better about their participation than participants forced to take part because they enrolled in a course. Positive feelings and attempts to “help,” on the other hand, may be more difficult to eliminate. If an experiment generates considerable concern with self-presentation, it may be desirable to add that factor to the theoretical foundation of the design.

For instance, a decision-making model developed by [Camilleri and Berger \(1967\)](#) included three sources of utilities in a situation: getting the right answer; pleasing the experimenter; and pleasing one’s partner. It was probably not possible to eliminate the experimenter effect here—the experiment was about getting right answers to problems, after all—but if effects can be conceptualized and measured in the situation, they become part of the theoretical foundation of the work.

Expectancy effects may also occur because experimenters may, unconsciously, treat participants differently based on experimental conditions. Because such differences are unconscious, often the experimenter is unaware of the changes in his or her behavior and thus it is difficult to control it. However, there are ways to reduce these effects. One way is to increase the number of experimenters. Some people will alter their behavior more than others; because it is difficult to detect when it is happening, one cannot simply eliminate those who do it more. By adding experimenters, the likelihood of this type of experimenter effect decreases. Of course, good training and observation of experimenters is also helpful in reducing this effect.

When possible, it is also helpful to reduce experimenter–participant contact. Technologies in place today can aid in that goal. Presenting the cover story via videotape or digital recording still provides the participants with something compelling (more so than does reading on paper or a computer screen), but it eliminates any condition-to-condition variability, such as that from experimenter fatigue, varied behaviors, and responses to participants. Videotaped instructions allow the experimenter to replicate exactly the information participants receive that is not varied across conditions, while also allowing him or her to edit in segments that do vary by condition. Data collection and measurement of the dependent variable by computer also help to eliminate experimenter effects.

Let us return to the extended example regarding “the gender experiment.” To reduce experimenter effects as much as possible, I did a number of things. The dependent variable data were collected electronically by computer and were not participant to observer effects. I limited experimenter–participant interaction by presenting the instructions, which were consistent across conditions except for short segments that related to the independent variables, via videotape.

In addition, I used a total of nine graduate and undergraduate student experimenters, all of whom underwent extensive training, to conduct the experimental sessions. The sessions included extensive interviews after the dependent variable data were collected; these interviews included debriefing about the deceptions presented in the cover story. These interviews were audio taped. When experimenters first began working on the study, I listened to a number of these audiotapes for each experimenter and provided feedback on reducing experimenter effects. Later, I randomly selected tapes for review to ensure continued control.

## **VI SUMMARY**

With this chapter, I hope I have provided both the novice and the experienced experimenter with some useful suggestions. By paying careful attention to the many details involved in the design of an experiment, we can be more sure that the data that result will be useful and meaningful. Whether you are planning your first experiment or your 20th experiment, the devil truly does hide in the details.

By paying careful attention from theory to hypotheses to variables to manipulations, you can increase the strength of an experimental design. Use of a standard protocol can increase your comparability to the results of others and can provide you with a vetted situation. You can strengthen the design further by considering the participants in your population (Chapter 5), through thoughtful use of technology, and with thorough training of the experimental staff (Chapter 4). Of course, no design is ever perfect, but the use of manipulation checks, pretests, pilot tests, and power analyses can help improve upon even the most well-thought-out experiments. Careful attention must not cease when the design is in place or even after the elements have been pretested; good experiments are an exercise in vigilance.

## ACKNOWLEDGMENTS

I thank the editors of this volume for their improvements to this chapter. This work was supported in part by National Science Foundation Grant SES 0317985.

## REFERENCES

- Camilleri, S. F., & Berger, J. (1967). Decision making and social influence: A model and an experimental test. *Sociometry*, 30, 365–378.
- Hunsberger, B., & Ennis, J. (1982). Experimenter effects in studies of religious attitudes. *Journal for the Scientific Study of Religion*, 21, 131–137.
- Keppel, G. (1991). *Design and analysis: A researcher's handbook*. Englewood Cliffs, NJ: Prentice Hall.
- Rashotte, L. S. (2006). Controlling and transferring the status effects of gender. In Paper presented at the meetings of the International Society of Political Psychology, Barcelona, Spain.
- Rashotte, L. S., Webster, M., Jr., & Whitmeyer, J. (2005). Pre-testing experimental instructions. *Sociological Methodology*, 35, 151–175.
- Willer, D., Walker, H., Markovsky, B., Willer, R., Lovaglia, M., Thye, S., et al. (2002). Network exchange theory. In J. Berger, & M. Zelditch, Jr., (Eds.), *New directions in contemporary sociological theory* (pp. 109–144). Lanham, MD: Rowman & Littlefield.
- Winer, G. A., Makowski, D., Alpert, H., & Collins, J. (1988). An analysis of experimenter effects on responses to a sex questionnaire. *Archives of Sexual Behavior*, 17, 257–263.

## Chapter 7

# Common Problems and Solutions in Experiments

Kathy J. Kuipers

*University of Montana, Missoula, Montana*

Stuart J. Hysom

*Texas A&M University, College Station, Texas*

### I INTRODUCTION

Typically, the goal of experimental research conducted in a controlled setting is to test theory or construct theoretical explanations. Independent variables are manipulated or introduced, and all other variables (extraneous) are carefully controlled in order for the experimenter to measure the dependent variable and make conclusions about how the variables are related. With attention to design details such as the introduction of variables, the measurement of dependent variables, and random assignment of subjects into control and experimental groups, what could go wrong? In this chapter, we suggest that there is much that experimenters must anticipate in making decisions about experimental design issues, maintaining the constancy of conditions across groups, and avoiding many of the common problems in conducting laboratory research. We discuss common problems and issues and offer some suggestions taken from experimental practice that may provide guidance for anyone designing and implementing an experiment.

To do this, we draw on our own experiences and training as researchers and experimentalists. While the problems on which we focus come from those training and research experiences within a university setting and our examples therefore are drawn from the university setting as well, many of these problems also may be encountered in other research settings with other types of subject pools.

### II RELATIONS WITH THE LARGER DEPARTMENT OR PROGRAM

We begin our discussion with some important considerations for setting up an experimental lab facility, designing space, and deciding on access issues. While initially a researcher may be grateful for any space that is quiet in which to

conduct experiments, unless that space is clearly designated for experimental research, it is difficult to control extraneous variables across conditions. When others have access to the space, the room may change over time, so the first subjects in an experimental run will participate in a room that is very different from the room in which the final subjects participate. An old poster (e.g., a political poster) or a stack of books may lead subjects to make assumptions about the hypotheses being tested or activate irrelevant emotional responses.

All such effects are undesirable. Laboratory space should be clearly dedicated for research and should be separate from the comings and goings of academic departments and programs. The space should be visually and spatially neutral, and external noise should be eliminated.

Experimenting, scheduling, and other laboratory jobs must be done by people designated for those positions. It might be tempting to use university secretarial or administrative help to schedule or refer arriving participants to a laboratory room, but that is not a good idea. For one thing, it would be an additional burden on a receptionist. Also, someone who is not a member of the research team cannot be counted on to do the job as it should be done—that is, avoiding unnecessary cues and other information and treating every person who arrives in almost exactly the same manner. Not having dedicated personnel handling all arrangements detracts from the seriousness, legitimacy, and importance of the research itself. Participants may view their participation as just an extension of their coursework, and experimenters risk subjects' task orientation and focus when their research is not separated from their teaching.

The experimental treatment begins when subjects first are contacted and consequently everything should be held constant except for the experimental independent variables. This means that much of normal interaction, such as greeting, asking how someone feels, and commenting on the weather or on sports, is inappropriate because it cannot be done uniformly for every experimental subject. Whereas an experimenter who receives subjects can learn not to engage in small talk, most departmental receptionists will not avoid it. The problem is to treat everyone exactly the same except for the independent variables, and that treatment starts with the first time anyone interacts with a subject in recruiting, scheduling, and arriving at the laboratory.

Because pay may be a part of the manipulation in an experiment (e.g., in a status-construction experiment in which pay level is a part of the status manipulation), another potential problem for experimenters is in how the payment will be distributed to subjects. Of course, careful bookkeeping records are required for the distribution of subject monies, and many complications can be avoided by handling the pay within the postexperimental interview. Pay also may be an important part of the debriefing. Typically, pay is distributed during the postexperimental debriefing when subjects are told about the purpose of the experiment, given an explanation for any deceptions, and have an opportunity to ask questions. The manner in which the debriefing is conducted will have an effect on how subjects feel about their participation (which may involve being

deceived), and receiving their pay at this time, rather than later, will make the experience more positive. But rules for payment are not universal. For example, in many social dilemma studies, subjects are told that they will be paid in private so that their decisions are not made public during the experiment. In such cases, subjects can be paid one at a time as they leave.

Equipment and facilities are often sources of problems encountered in a laboratory. While most research involves the use of computers for data storage, analysis, or reporting, laboratory equipment is unique in that it must be operating correctly while subjects are present. Conduct during the experiment, from the arrival of the subject until the completion of the debriefing, can be seen as a performance with the subjects as an audience. (In Chapter 4, Shelly develops the theatrical analogy.) Equipment failure creates an impression of incompetence and unscientific work. Also, because all features of the experiment must be controlled and constant (with the exception of treatment variables) in order not to contaminate the experimental results, extraneous variables such as broken or misbehaving equipment must be eliminated. It is therefore important to check equipment thoroughly before the experiment begins, monitor it in pretests, and make periodic checks during the experimental run for any malfunctions.

Some labs are able to pay for technical support for maintenance and service to equipment. While this option can be a great help, particularly when subjects are tightly scheduled and repairs need to be done quickly, it is important that the relationship with information technologies staff is explicitly discussed. If staff is involved in maintenance, it is important to discuss scheduling so that everyone knows when they may or may not enter the subject rooms. If someone stumbles in while a study is being conducted, his or her presence would have an effect on the measurement of variables in a way that is impossible to determine and the contaminated data would be unusable. In addition to information technologies support, other equipment in the lab may also malfunction. Experimenters have found it wise to buy tools and a toolbox in advance for quick fixes to furniture, tacking up cords, and making minor repairs to equipment.

To avoid most of the problems with unwanted visitors, especially while an experiment is under way, the experimental lab should be locked at all times. In our experimental labs, we find it helpful to post signs clearly identifying the area. (This also provides research legitimacy from the subject's point of view.) We also post a sign requesting that no one enter the subject rooms, including custodial personnel for cleaning and services. Unless an experimenter can be certain that custodians only work at times when data are not being collected, and that they will never move anything in the laboratory—which usually cannot be guaranteed—it is best to ask them not to enter and to do his or her own routine dusting and emptying of wastebaskets.

Locking the lab and any file cabinets and computers in which data are stored is also important for protecting the confidentiality of participants. Institutional review boards will require that all records be kept in a secure location, as Hegtvedt describes in Chapter 2. Signed consent forms should be stored

separately from other experimental data in locked file cabinets or cupboards. Locking the lab also will minimize the possibility of items or records being misplaced, moved, or taken by someone who is not a part of the research project. Potential subjects and payment money will be wasted if subjects show up for a study and experimenters are unable to run the experiment because any of the needed items are missing.

Because experimental manipulations are often complicated, requiring several forms and questionnaires, and because experimenters are less likely to bias subjects when they run more than one condition during a time period, it is easy for mix-ups in materials to occur. An experimenter may grab the wrong form for a condition if everything is not clearly labeled and stored. To minimize such mistakes, we post clear instructions and orders of procedures in the laboratory control room. We also label and separate forms and documents for the different conditions even if some of the same are used for more than one condition. Many mix-ups can be avoided by keeping materials for each condition together and forms and questionnaires clearly labeled.

### III EXPERIMENTAL MANIPULATIONS AND DECEPTION

In laboratory research, manipulation refers to the construction of events or information in a controlled environment. Most often, experimenters manipulate the independent variable by how, where, or when it is introduced into the experimental situation or by the level of its introduction. They also may manipulate information provided to the subjects about the task, their partners, and even the subjects themselves. This manipulation may involve the use of deception—deceiving the subject about the true nature of the study or its hypotheses—or deception may be used when other aspects of the experiment are manipulated in order to introduce scope conditions (that must be met in order for the theory to apply) or to control other extraneous variables that may have an unwanted effect on the dependent variable.

#### A Deception or Nondeception?

The first decision regarding deception is whether or not to use it. While we may mislead or deceive subjects through the presentation of partial information about the hypotheses or research question, the intentional misrepresentation of tasks or actors is true deception ([Sell, 2008](#)). One way in which deception often is used in experiments is to prevent subjects from learning the true hypotheses. Once subjects are aware of the hypotheses, they may not behave as they would without that awareness. They may consciously agree with the predictions and try to “help” experimenters, or they may disagree with the predictions and try to show experimenters how they are wrong. In other cases, subjects simply may be unconscious of the influences that such knowledge has on them, although its influence is well documented in experimental research (e.g., see [Orne 1962, 1969](#);

(Roethlisberger & Dickson 1939; Rosenthal, 1967, 1969, 1976). Subjects may pick up clues about the hypotheses or goals of the experiment, and this may influence a change in their behavior to go along with what they think is demanded of them. Subjects create demand characteristics with knowledge or guesses about the expectations that researchers have for their behavior. Experimenters should be keenly aware of how demand characteristics may form. One way to minimize the effects of demand characteristics is to make sure that subjects are unaware of hypotheses and of experimental manipulations.

Ethically, experimenters are obliged to limit deception unless unavoidable (see ethical codes for various professional associations, such as the [American Sociological Association's Code of Ethics \(1999\)](#)). If it is possible to conduct the experiment without deceiving the participants, that is always preferable. It is acceptable to use deception only if it has a specific purpose in the study, if it does no permanent harm, and if the benefits from participation outweigh any negative effects. Deception should be used only to the minimal degree necessary. So, while experimenters may not reveal their complete hypotheses, they will try to give information that is not completely untrue to subjects about what subjects will be doing. It is always best to reveal everyone's actual role in the experiment and their purpose, if possible. The main exception to this rule is designs requiring the use of confederates.

Researchers must, of course, obtain informed consent from participants, and this is underscored when deception is used. Although some deception is acceptable, researchers must never misrepresent any potential risks to subjects. If there are risks or discomforts that are likely to affect a participant's willingness to continue with the study, those should be revealed. An important part of the postexperimental interview must include a debriefing in which the participant is informed of exactly how he or she was deceived, given an explanation for why the deception was necessary, and assured that his or her behavior was consistent with that of others in the study. Experimenters explain to their subjects that they were fooled by deception because the experimenters went to a great deal of trouble to construct the measures and tests so that nearly everyone believed what they were told. After deception has been explained to subjects, the debriefing session also should include opportunities for participants to express concerns and ask questions in a safe environment.

Before participation in any experiment, the subject should give his or her informed consent. (For a more complete discussion, see Hegtvedt, Chapter 2.) Information on the consent form will reveal or clarify how many subjects are involved, what the study generally entails, and how responses are being recorded. Subjects give their permission for their participation in the study by adding their signatures to the forms. One form is kept for the experimental records and the subject receives the other form.

If deception is a part of the study, there are a variety of acceptable ways in which to present it. Subjects will want to know a little bit about the study before they begin, and for this reason, a cover story is typically given. The cover story

is an explanation of the purpose or nature of the research, and it accomplishes several things. First, it assures the subject and makes him or her more comfortable in an unnatural situation. The cover story also should arouse some interest for the subject whose attention to the task and the various elements of the study is essential. If the subject is not paying attention, the outcomes will be irrelevant. In addition, the cover story will lessen a subject's natural preoccupation with the hypothesis and the true purpose of a study by providing an explanation for the setting in which he or she will be asked to participate.

Subjects may be intentionally misled through written or verbal instructions. In many experiments, subjects are told that the task in which they are about to participate measures a particular ability. For example, there is a tradition in expectation states research to use tasks that are intentionally ambiguous in order to prevent subjects from having special information or skills that might make them experts at such tasks and influence the outcomes. At the same time, they are told that there really are right and wrong answers (when there are none) in order for them to be task oriented. Subjects also may be told something about themselves that is not true. In some trust experiments, subjects are told that they will work with a partner who is trusted by the project director—trusted to make the best final decisions and to distribute pay fairly, for example. Subjects also may be deceived through the actions of others, either the experimenter's behavior or that of a confederate. Later in this section, we discuss the use of confederates in laboratory experiments, but experimenters also will be acting to create a certain impression. The design of an experiment may require false treatments or dependent variable measures. In status-construction experiments ([Ridgeway, Boyle, Kuipers, & Robinson, 1998](#)), for example, in order to create a nominal characteristic that carries no prior meaning for each subject, subjects complete a "personal response style" test (a task adapted from social identity studies; [Tajfel, Billig, Bundy, & Flament, 1971](#)) and are told that this information will determine whether they belong to a group called S2s or to another group called Q2s. In actuality, there is no such characteristic and tests simply ask subjects to distinguish between reproductions of paintings by Klee and Kandinsky and indicate their preferences. Once the tests are "graded," a subject is told his or her "personal response style" type—making him or her dissimilar from his or her partner. Deception also may take place through manipulation of features of the setting, such as arranging furniture to structure interaction or separating outcome measures into what appears to be two separate studies to reduce suspicion about the true goals of the study.

So, how do experimenters make choices about deception? Not only are ethical concerns at stake when using deception but also if the deception involves many aspects of a study, there is a greater chance for the deception to be discovered by the subjects. It may be best to give subjects just enough information to draw conclusions but not to give additional false information if it can be avoided. Rather than telling subjects something that is untrue, experimenters can construct situations that will allow subjects to draw their own conclusions.

Those conclusions may not be accurate, but compared with falsehoods and lies, they are easier to correct in a debriefing. For example, in the status-construction experiment mentioned previously (Ridgeway et al., 1998), subjects are asked to fill out a background information sheet. After the information is collected and evaluated and another test is taken, experimenters tell subjects how much pay they will receive “based on the information the laboratory has about [them] and the other participants.” Subjects do not know what information is being used, but by implication, some of that information comes from the background information sheet. In the debriefing, great care is taken to assure subjects that the pay they were assigned had nothing to do with the background information that they supplied.

## B Strength of the Manipulations

In designing experiments, decisions must be made about the strength of the treatment variables and how, where, and when they will be introduced; it is important to consider how strong the presence of a variable will be. In general, we desire the most conservative test possible for our hypotheses, so experimenters introduce the independent or treatment variable in such a way as to allow the subjects to have choices in how they will behave. Laboratory interactions are structured to make alternative behaviors possible—not just those that are predicted but also behaviors that would disconfirm hypotheses. When participants are allowed to make choices in their behavior and in their responses, there is a clear measure of the relationship between treatment and outcome decisions under controlled conditions.

On the other hand, experimenters must make sure that a manipulation is understood or interpreted correctly—not necessarily in a conscious sense—but subjects must be aware of experiencing treatments that the experimenter predicts will influence their responses. The tendency is to load up on the indicators of the treatment in order to make sure that they are not missed. In the status-construction experiments, subjects were told about their pay levels at several points in the study, they read about their pay levels on the computer and on a payment form, they were asked to write their pay levels on several different forms, and they were later interviewed about pay levels on questionnaires and in person. Not only was the information repeated to make sure that subjects were aware of their pay and their partners’ pay but also several manipulation checks were conducted to make sure that subjects understood the differences and how they compared with others.

Manipulation checks are conducted to assess whether an independent variable is experienced or interpreted in the way that the experimenter intends it to be and whether the scope conditions (discussed later in this chapter) have been created. Checks may be conducted immediately after a variable is introduced, when subjects will be more likely to recall its presence. In some experiments, these manipulation checks indicate if aspects of the experiment are well enough

understood by the subjects for the experiment to proceed. For example, in many social dilemma experiments, manipulation checks ensure that subjects understand the payoff structure of the experiment. In other experiments, however, manipulation checks anywhere but at the end of the study would arouse suspicion or otherwise distract subjects from the task at hand. Of course, this especially would be the case for experiments involving deception. It is safer to place manipulation checks at the end of the experiment in the postexperiment interview. Some argue that subjects may have forgotten the manipulation by this time. If a variable is manipulated as strongly as it should be for good experimental design, however, subjects will recall it, and the risk of influencing the dependent variable measures is reduced.

A common term for actors who work for the experimenter and pose as other subjects or as bystanders is “confederates.” Confederates deliberately mislead subjects and typically participate in the introduction of an independent variable, modification of the decision process, or control of the setting within which predicted outcomes are expected to take place. The use of confederates presents ethical problems related to deception, as discussed previously. In addition, depending on how confederates are used, they may introduce more extraneous variables into the experimental setting if their introduction is not carefully scripted and controlled. Because each participant in all experimental conditions must have exactly the same experience under the same controlled conditions, it is vital that confederates behave in almost exactly the same way with each subject.

In the original status-construction experiment, subjects worked on a task with a partner in a doubly dissimilar encounter. That is, the encounters took place between actors who differed in two ways: on a nominal characteristic and in terms of the level of resources each would receive. Subjects were scheduled in groups of four and had an opportunity to work with a series of partners about whom they received additional information. Researchers immediately discovered problems in pretests with this design because subjects were able to view and take into account a variety of characteristics about each other that were beyond the control of the experimenters. These characteristics, such as dress, facial and physical features, voice, and congeniality (real or inferred), combined with the doubly dissimilar characteristics to influence outcomes—status beliefs and observable power and prestige in the group interaction. It was impossible to ascertain the effects of the doubly dissimilar encounters by themselves.

Instead, to control information about dress, physical features, voice, and congeniality that subjects received about their partners, confederates were used in the final experimental design. Each subject worked with two confederates at two different times, interacting through an audio connection for verbal interaction. Not only did the use of confederates reduce much of the contamination or extra noise that interfered with the doubly dissimilar encounters but also it allowed us to control the behavior of the partner, an important cue for the determination of status construction.

In many social psychological studies, interaction is a necessary component of the process investigated. In such situations, if the interaction must be face-to-face, the use of well-scripted confederates will provide many of the necessary controls for extraneous variables. In addition, the easiest way to control confederate behavior is to restrict the type of interaction he or she has with the subject. For example, in the experiment discussed previously, subjects were allowed only audio communication with their partners when discussing their decisions in order to cut down on the other cues that are provided in a face-to-face setting.

Often, subjects work with a partner on a task. If it is not necessary that the subject see or hear the confederate with whom he or she works, the confederate/partner may communicate with the subject electronically through computer interaction. In that case, the confederate/partner may be programmed into the computer and all interaction apparently generated by the confederate/partner actually will be generated by the computer. It is important for a subject to have some indication that his or her partner is, indeed, a real person because computer interaction is not rare. Fortunately, it is now possible to locate software that allows an experimenter to program in specific responses on the part of a confederate/partner to keystrokes from a subject.

If subjects must see and/or hear the other people with whom they work on a collective task, again, technology can provide some solutions. Subjects may view on a monitor partners who are presumably working in another room. These partners could be previously audio or video recorded so that their behavior and characteristics are exactly replicated for each case. (Later in the chapter, we discuss other techniques and considerations in setting up an experiment using confederates.)

In most of our laboratory experiments in sociology, we are interested in small group interaction. By small group, we usually mean from 2 to 20 participants—only enough for the participants to have face-to-face interaction. One consideration in designing experiments must be the size of the group. In general, the larger the group, the more possibilities there are for extraneous variables to affect aspects of the experiment. If groups are mostly or entirely made up of naive subjects, there may be scheduling problems. Problems with scheduling only one subject and getting him or her to show up in the right place at the right time are magnified by the number of subjects in the group. It will require always scheduling additional participants as substitutes and, of course, paying them even if they are not needed. This increases the cost of an experiment.

There is no formula for deciding the exact size of a group, unless size is a theoretical variable, as it is in some network experiments. Experimenters usually try to limit the group size to the minimum number of participants possible without endangering the manipulations. It is most important to consider the theoretical concepts under study. For studies focused on interaction, it may be possible to observe most features of groups, such as subordination, in a dyad. For example, we can observe status differentiation with just two individuals working with each other even though we may conceptualize that differentiation

in much larger groups. At the same time, the dynamics of group interaction differs greatly when going from a two-person group to a three-person group, and those features may be theoretically important to capture. Some experiments may require examining organizational aspects of groups, and in this case, the size is dictated by the structure of the organization in question. [Johnson \(1994\)](#), for example, creates organizational structure with only three people. Other experiments may require large groups per se. Recent investigations of such large groups have used computer networks to facilitate group size.

In addition, if more than one confederate must be used, experimenters will try to limit the interaction as suggested previously. If participants must interact with each other, again, the nature of that interaction should be limited as much as possible.

## IV EXPERIMENTAL DESIGN ISSUES

We often have referred to the design of an experiment in discussing manipulations. Manipulations are not the only place where problems can occur, however. Experimenters need to consider the length of the experiment as it influences subjects' and experimenters' fatigue, the emotional stress experienced by subjects due to an unfamiliar experimental environment, and the level of involvement for subjects in the task.

### A Length of the Experiment

It is important that the experiment not be too long for a variety of reasons. Subjects may become fatigued or sleepy. They may become bored and lose interest in the task and in performing well. They may get hungry or have other needs that necessitate the interruption of the experiment. Of course, if an experiment is interrupted, that subject's experience is no longer the same as that of the other participants and the controlled setting is contaminated.

It is also more difficult to maintain deception over longer periods of time. The more subjects interact with a fictitious partner or under ambiguous circumstances, the more opportunities they have to become suspicious of the manipulations. They may grow suspicious over time. Subjects also have more opportunities to spot deception, to wonder about features that have been glossed over, or to be offended by someone's behavior. Of course, if an experiment is boring, requires many repetitions of a task, or is otherwise not very engaging, the likelihood for subjects to lose interest and focus will increase.

Therefore, experienced researchers try to keep their experimental tasks short and to limit experimental trials. If experiments must run longer because they require several manipulations with follow-up tasks or retests, subjects should be allowed to get up, change positions, move from one room to another, or to do different tasks.

## B Stress and Discomfort

Related to the concern over time limits in laboratory experiments are stresses and discomfort levels for participants. Stress may occur as a result of emotional strain (embarrassing activities or information that negatively affects self-esteem), physical strain (long or fatiguing tasks), and mental strain (repetitious tasks or instructions that are complicated or difficult to remember).

To avoid potential problems, researchers should anticipate that subjects may be made uncomfortable by offensive words or situations. Pretesting or pilot testing of all materials and manipulations should be conducted well before the experiment is scheduled to begin. Pilot subjects should be run through the experiment as if they were real subjects. Afterwards, the researcher should interview the pilot subjects, looking for any aspects of the experimental design that may cause subjects discomfort.

## C Involving the Participants

Often, experimenters ask their subjects to perform a task that may or may not be related to the actual measure of the dependent variable. The involvement of the participants in the task is essential for experiments in several ways. Frequently, a scope condition for an experiment is that participants are oriented toward a valued task—they are not distracted and they want to do it well. In addition, subjects often must work with one or several partners on a task in order for interaction to be measured. To ensure meaningful results, the task itself must be somewhat engaging. At the same time, in order to control familiarity with the task, expertise of subjects, and perceptions of self in regard to task performance, the task should not be one that subjects have done before (unless that is a condition of the experimental hypotheses). For example, [Ridgeway et al. \(1998\)](#) asked subjects to match early language terms, and expectation states researchers have asked subjects to assess contrasting black and white squares for which pattern had the most white color ([Berger, Fisek, Norman, & Zelditch, 1977](#)). These tasks are challenging and they hold subjects' interests without becoming too tedious. Subjects are often intrigued and curious about their abilities on such tasks and are anxious to find out how well they performed.

Another way that experimenters are able to ensure task focus is to offer a reward for doing the task well. In many social dilemma studies, pay for subjects' particular contributions is given, and in fact, it is often the dependent variable. In status-construction studies ([Ridgeway et al., 1998](#)), we offered to pay the group with the best score a bonus of \$50 to be divided between the participants. It is often common to offer each group a bonus if they do particularly well. This possibility for an additional monetary reward oriented the subjects to the task.

Of course, experimenters should check for involvement, along with other conditions essential for the hypotheses, in the postexperiment interview. Experimenters ask subjects what they thought of the task, if they tried to do their best, and if they have any questions about it.

## V RUNNING EXPERIMENTS USING CONFEDERATES

Once a decision has been made to use one or more confederates in an experimental design, an experimenter will need to consider who will play that role. Confederates posing as subjects should match characteristics of the subject pool as closely as possible. To accomplish that, it is best to select confederates from the original subject pool. Professional actors are not necessary; however, confederates who are able to act as if they are subjects without arousing suspicion are essential.

In fact, actors present their own problems, which you might not expect. Although they can be very good at portraying a role, that is not necessarily what is needed in experimentation. Part of a good theatrical performance is individuality—portraying a role with enough spontaneity that it creates a memorable “person.” Individuality and memorability, of course, are inappropriate in laboratory settings. Unless an experimenter plans to hire a director, he or she may be well advised to try training nonprofessional students to serve as confederates.

In addition, confederates should not be people who might be known to the subjects before they participate in the study. In Ridgeway et al.’s (1998) status-construction studies at Stanford, a large number of confederates were used. They were hired from the student body (because our subjects were undergraduate students), and although subjects would not actually see them, confederates were chosen who resembled and sounded like typical students (male and female). Confederates were told to arrive early, but their physical resemblance to the subjects reduced any chance for contamination created by a quick glimpse from an uneasy subject. Confederates were also told to dress like typical undergraduates, unassuming, and without any props that might affiliate them with specific student subcultures. The goal with confederates (as it is with the physical layout of the laboratory) is to avoid any unusual or memorable features.

Experimenters will need to determine how many confederates to employ. Obviously, consistency across experimental groups is essential, so one might imagine that one well-trained confederate could ensure that each subject’s encounter will be exactly the same. The opposing argument is that if confederates are well trained, more than one may be advantageous. Multiple confederates may be advantageous because of problems of scheduling, confederate fatigue, and illnesses. When a research team includes multiple confederates, it is important to assess confederate performance. Confederates can be compared in their actual performance, in assessments on session reports and postexperiment interviews, and through evaluations on post-task questionnaires. If confederates are well trained and if the script (discussed later) is functioning effectively, results should be similar and consistent. Of course, the demands of the confederate role, complicated scripts, and elaborate procedures all will influence the decision about how many confederates to use. If the role is demanding and requires considerable nonscripted interaction with subjects, then fewer well-trained confederates will make fewer errors.

## A Training Confederates

Thorough training is essential. Experts suggest beginning with a manual for confederates with the following sections: the importance of consistency and control in experimental research; the role that the confederate will play; what the confederate should say and do; what the confederate should *not* say and do; and what may be said and done in response to the subjects' behaviors and comments. For example, in the status-construction experiment, confederates' instructions included the following:

*Please be careful about mistakes, note them on your session report (at the end of the experiment) if any occur, and make sure that you follow the script and order of procedures EXACTLY.... Remember that it's not o.k. to ad lib your responses. Everyone MUST BE SURE TO SAY EXACTLY WHAT'S ON THE SCRIPT, at least initially when you first announce your choices. This is how we make sure that all subjects are having essentially the same experience. Of course, once you make your initial proclamation, you will have to do some acting. And you are the best judge of how to do that—different subjects require different types of responses for you to communicate your deferentialness or nondeferentialness.... Use the comments posted on the sheets on the wall above your desk.*

Confederates should always work from a very clear script with all dialogue included. Of course, the less a confederate is expected to say, the fewer possibilities for error. The behavior and speech of a confederate also must be as realistic and normal sounding as possible. If some of the information that a subject will receive about a confederate can be controlled, such as having the confederate in another room where his or her behavior will not be observed, it is advisable. Then, spoken interaction is the only arena for errors or miscommunications. In the status-construction experiments, what the confederates initially said was very carefully scripted. In addition, confederates were given lists of possible comments to make depending on how the subject responded.

Practice sessions are critical. The practice sessions can be recorded in order for the experimental research team and the confederates to go over the recordings to critique them. This is also an opportunity for the confederates to comment on any problems they see with procedures or with scripts.

## B Managing and Assessing Confederates

At the end of each experimental session, the confederate should fill out a session report giving comments and impressions about how the session went, his or her performance as a confederate, and whether any suspicion on the part of the subject was detected. These reports are valuable not only before the actual experiment begins but also to evaluate confederate performance in pretests and throughout the study as confederates become more confident (and, perhaps, less vigilant) in performing their roles. In the status-construction experiment,

the confederate session reports included questions about whether the subject was cooperative, task oriented, seemed to try to do his or her best, seemed to believe that the confederate was a real subject, and took the task seriously. Confederates also were asked if they recognized the subject (since that would influence the acting), if they detected any characteristics of the subject other than what they had been told by the experimenter, and whether they had any technical problems.

Confederates benefit from monitoring—observation and listening in on their acting—and frequent meetings about problems encountered by both confederates and experimenters. Audio recording verbal interactions and reviewing those recordings after the experiment allows the experimenter to focus on the subject while the experiment is being run and attend to confederate performance later.

Confederate performance can also be assessed in a post-task questionnaire and in postexperimental interviews with subjects, asking about their perceptions of the people with whom they worked. Subjects can be asked similar questions about their own performance and about the confederates' behaviors: evaluating task orientation; team orientation; and believability of the manipulations.

## VI DEVELOPING PROCEDURES

Experimental procedures must be developed such that every interaction between laboratory personnel and research participants is replicable. This requires careful development of word-by-word scripts, complete with blocking (i.e., stage directions) and intonation information for the entire experimental session and all scheduling and reminder phone calls. Developing these scripts is a time-consuming process. A common error when planning experiments is to schedule too little time for the development of these materials. The best way to save time and other resources here is, of course, to locate and use procedures that already exist and that have been used successfully by others. Using an existing design that can be modified for a new experimental purpose is far less likely to cause unanticipated problems than designing an entirely new experiment.

When constructing a new script, it is useful to break the task into smaller pieces. Distinguish between those parts that are standardized across conditions (e.g., phone calls, participant arrivals and greetings, securing informed consent, paying participants, task instructions, and measurement) and those parts that are related to the manipulation of the independent variable, which differ by condition. It is perhaps easiest to work first on those parts that are shared across conditions and then move on to the manipulation.

Rashotte, Webster, and Whitmeyer (2005) suggest beginning with a list of *scope conditions* and *initial conditions* for the experiment. Scope conditions abstractly define the domain, or type of social situation, to which a theory is intended to apply (Walker & Cohen, 1985). Initial conditions describe concretely

what will be done in the laboratory to instantiate the scope conditions and other features of the situation such as the type of interaction that will be used (Rashotte et al., 2005).

Expectation states theory's scope conditions state that actors are: (1) collectively oriented (2) toward completing a valued task (described by Berger in Chapter 12). In the standard expectation states setting (also described by Berger in Chapter 12), initial conditions instantiate a collective task by asking participants to work with a partner (who may be simulated by a computer) on trials of one or more tasks, each of which is said to involve a different newly discovered ability (e.g., contrast sensitivity), and which are said to be unrelated to other well-known abilities. Instructions usually include a presentation of "scoring standards" about how previous teams of similar students have scored on the test so that subjects have a point of reference from which to interpret their own scores relative to those of others, especially those with whom they will be interacting. Collective orientation is encouraged by telling participants that individuals working with others often make better decisions than they would when working alone, and that the study at hand is interested in learning about "just this type of situation."

Thus, the list of initial conditions would include the following information being delivered to the participant: (1) that individuals working together often make better decisions than they would working alone; (2) that several newly discovered abilities are being studied, and that they are important and interesting because they are known to be unrelated to other more common abilities; (3) that each task problem has a correct answer, and that individuals with high levels of the ability typically get more correct final choices than do individuals with low levels of the ability; (4) instructions for using equipment (to indicate initial and final choices and receive feedback on partner's initial choice); and (5) scoring standards. In the expectation states protocol, standardized instructions have been developed for delivering this information (see Berger, Chapter 12; Berger & Zelditch, 1977; Cook, Cronkite, & Wagner, 1974; Troyer, 2001, 2007).

When developing new procedures, simple phrasing is best so that each important idea is clearly and deliberately communicated (although initial phrasing may require modifications based on information gained from analyses of pretest data; see later discussion). Other general guidelines include expressing one idea in each sentence and avoiding complex grammatical constructions. If the procedures do not instantiate scope conditions for the participant from the participant's point of view, then his or her behavior will be irrelevant to assessing derived hypotheses. Therefore, key aspects of the instructions are typically repeated at least three times, using the same or slightly different phrasing each time. Because laboratory experiments aim to create a simplified, and therefore "artificial," social situation, it is not necessarily important that the instructions sound natural or conversational. Rather, what is most important is that, from the perspective of participants, the procedures instantiate

scope conditions. The extent to which a set of procedures do, in fact, instantiate initial conditions can only be assessed through pretesting the procedures, the task we consider next.

## VII PRETESTING

Once an initial script (and an initial video recording, if one is being used; see later discussion) is complete, it should be assessed during a pretesting phase. The overall goal in pretesting is to learn how well the procedures create the required initial conditions for study participants and whether the independent variable manipulation is adequate. Some specific questions to address during pretesting might include the following: Did the participants treat the situation seriously? Did they understand what they are being asked to do? Did they notice and believe manipulations?

Even when the experiment is only a slight modification of a set of standardized procedures, it is desirable to run initial “practice” sessions. For these, it is usually most straightforward to run the simplest condition (often the control condition) first and use graduate students or undergraduate students involved in conducting the research as stand-ins for participants. Such an initial run-through will allow quick identification of problems in stage directions (e.g., failure to note in the script that research assistants must deliver to the participant a writing implement along with any paper-and-pencil instrument) and timing (e.g., allocating 5 minutes for participants to complete a postsession questionnaire when the instrument actually takes 40 minutes to complete). Initial groups also help to assess the usability of data collection instrumentation in real-time and provide training opportunities for research assistants.

To assess other aspects of the instructions, pretest participants should be drawn from the same population as will be participants for the study itself. Manipulation checks (as discussed previously) provide important information for assessing design. For example, if information about the participant’s and partner’s scores is provided as part of a manipulation, a postsession questionnaire item might ask a respondent to recall and write down her and her partner’s scores. To assess collective and task orientation, an item might ask, “When working on the contrast sensitivity task, how important was it for you to get the *correct* final answer?” along with a 7-point anchored scale running from “extremely important” to “not at all important.” Collective orientation might be assessed using similar items asking how important were “getting the right answer,” “sticking with initial decisions during disagreement trials,” and “changing your initial choices just to agree with your partner when your initial choices were different.” Each would have anchored scales running from “extremely important” to “extremely unimportant.”

Multiple measures for each important concept are usually desirable. Questionnaire items are an economical way to gather information about how the participant understood and interpreted the situation, but to fully assess the

effectiveness of procedures, pretest participants usually are interviewed at some length immediately following the experimental session. When an investigator suspects that the situation did not, in fact, instantiate the desired initial conditions for a given pretest participant, follow-up questions should ask carefully about the participant's experience to learn as much as possible about what happened and how the participant experienced the situation.

In Hysom's (2004) dissertation research, for example, postsession interviews indicated that several participants in a low-status condition did not accept their low scores at a fictitious ability (meaning insight). These individuals did not appear to suspect that the ability, the task, or the scores were fictitious; rather, they appeared to discount the importance or meaning of having received a low score. One possible explanation for this was that these particular students, attending a select, private university, might have had little personal experience doing poorly in other (mostly academic) testing situations. It seemed possible that for these participants, the inconsistency of their low task score with previous scores on other tests might have led them to downplay its importance.

In a revised version of the script, the disassociation between meaning insight and academic performance was made explicit. Phrases were added, such as that the ability was also unrelated to scores on standardized exams such as the SAT, one's GPA, "or even past performance in a classroom situation." This information was reinforced multiple times. In subsequent pretesting of the revised script, no participants in the low-status condition said anything to indicate that they did not accept their scores. Procedural changes based on pretest feedback also usually are pretested in order to verify that goals of changes are met.

Postsession interviews with pretest participants are typically much longer than interviews with actual study participants. Orne (1969) recommends, for example, enlisting pretest participants as "co-investigators," informing them of the study's purpose, explaining in some detail what is being measured and how, and asking—from their perspective—how they made their decisions, arguing that such information is useful in assessing demand characteristics. Another assessment that can be made during pretesting is reactivity of participants to the procedures: does anything in the instructions offend or anger participants? If yes, procedures should be adjusted to avoid such reactions, just as they should be changed if, for example, pretest participants find the task instructions confusing or difficult to follow. Many participants are eager to discuss their experiences. Early nondirective questions can encourage this (e.g., "Well, what did you think?"), with later questions becoming more specific (e.g., "After you made your *initial* decision, and then saw your partner's answer, what did you do then?"). Care should be taken that no early questions inadvertently provide information to participants that may alter their responses to later questions.

The focus of postsession interviews changes once pretesting is complete. During pretesting, the function (beyond the essential debriefing and payment functions) is to assess the procedures so that they can be improved. Once a study commences, procedures must not be altered, so this function becomes

the determination of whether, for each given participant, scope and initial conditions were *not* met. If, for a given participant, the theory's scope conditions are not met, then predictions derived from the theory are not expected to apply. Therefore, the experimenter should exclude data from that individual because his or her responses are irrelevant to evaluating the predictions under test. Of course, excluding participants introduces a possibility of biasing results, and such decisions seem especially likely to be vulnerable to experimenter effects (discussed later). Therefore, the default decision is to include a participant unless a specific, pre-established reason for exclusion definitely exists.

Decision rules for exclusion are developed before any study sessions are run, are uniformly applied to participants, are conservative so that the subject is not easily excluded, and are explicit. A decision rule for “lacks collective task focus,” for example, might state: (1) that a participant must provide unambiguous information that she came to a conclusion that it was not necessary, important, and/or legitimate to take her partner’s initial choices into account (e.g., “I didn’t believe my partner was real, so I just ignored him”); and (2) that she therefore changed her behavior in some concrete way (e.g., “So I just chose answers by alternating between the left and right buttons”). In addition, exclusion decisions are made at the end of each interview. Such decisions may be reviewed by the experimenter later—for instance, by listening to a recording of the interview—but such second guessing suffers from a lack of the many nonverbal cues that are present during an interview (e.g., smiling, avoiding eye contact, or rolling one’s eyes).

Other important information can be gained during pretesting. One advantage of using a standardized protocol, for example, is that pretest results can be compared with past studies if comparable conditions are run.

Laboratory personnel who are authoritative, friendly, believable, and competent help ensure that study participants accept the information that is presented and follow directions. It is therefore important to train assistants carefully, before commencing a study, and stress that every interaction with every participant is “part of the study” and that these interactions, therefore, must be standardized. Experimental instructions are delivered at a relatively slow rate so that subjects understand each aspect of the study. Along these lines, it is important that all research assistants speak at the same rate (slower, more measured than everyday speech), emphasize the same words (bold typeface in a script can indicate words to be stressed), pause at the same places for the same duration (pauses are noted using boldface slashes), and maintain eye contact similarly. During training, modeling these behaviors multiple times is an effective way to develop standardization.

## VIII VIDEO RECORDING

Video recordings, from films to videotapes, have been used by experimental researchers for decades, both to record interaction or, more generally, to collect

behavioral data and to present stimulus material and instructions to participants (see [Dowrick & Biggs, 1983](#)). Recently, with easy-to-use computer software, digital cameras and recording devices, and multiple ways of creating and storing digital files, digital recordings are now preferred. Although digital recording equipment and computers are often easier to use, it is still advisable to have back-up equipment, cables, and digital storage space (if recording interaction) and multiple back-ups of instructional files (if using prerecorded instructions). Next, we discuss in some detail the recording of group discussions and then the production and testing of video instructions.

## A Recording Interaction

Poor-quality recordings can lead to many difficulties. Audio and lighting problems, for example, can make analysis of recorded interaction much more difficult because if a coder cannot understand what participants are saying, it is impossible to transcribe the audio recording accurately. If a participant pushes her chair back from a table, and thus out of the camera's frame, then her behavior is not visible, and so it also will be impossible to code. Bright fluorescent lighting can wash out contrast, so incandescent desk lamps should be considered to improve recording quality. Before running any groups, it is useful to test the quality of recording that results from different placement options for cameras, lights, and microphones. Verify also that any props provided to participants cannot easily be placed where they will block the camera. If facilities are not soundproof, outside noise may distract participants and may make the audio more difficult for coders to understand. Although newer models of most video cameras have very good microphones for audio recording, testing sound quality may determine that additional microphones are needed. Directional microphones are one good way to avoid picking up extraneous noise. Lapel microphones are good for capturing speech, but problems arise if participants touch or play with the microphones because the noise this produces in the audio recording can make analysis of the recording more difficult. When purchasing microphones, it is important to consider that there are different types with different pickup patterns.

When recording interaction such as a discussion group, it is usually necessary to have at least two cameras so that participants sitting across from one another at a table can be captured. [Johnson, Fasula, Hysom, and Khanna \(2006\)](#), for example, arranged three discussion group participants so that one sat on one side of a desk while the other two sat on the other side. One camera was mounted high on the wall behind the desk pointed at the two participants on one side, and another camera was mounted high on a wall across the room, pointed at the single person on the inside of the desk. Cables were routed inside the walls. Another option for discussion groups is to use a dedicated camera and lapel microphone for each participant. When multiple cameras are used to record discussion groups, a screen splitter is a useful piece of hardware. It takes

as input the video signals from two or more cameras and sends the images from each camera, arranged one in each half (horizontally or vertically) or quarter of the monitor screen and then directly onto the computer. This eliminates the need to synchronize multiple video files of a single group, postrecording, which is necessary for coding reactions of participants toward one another during interaction.

Camera position is determined most by the type of data that will be extracted from the recording, taking into account the physical facilities available. If a task requires participants to move about, dedicated cameras may not be possible. Cameras and microphones should be unobtrusive. Tripods take up a lot of space and are awkward to work with. Loose cables can be potential disasters. Consequently, mounted cameras on the wall and cleanly routed cables can be important for subject interaction. When arranging chairs and surfaces for discussion groups, it is useful to limit their possible movement. It is also important that laboratory assistants verify, for each experimental session, that all participants are in frame before taping begins and that microphones are turned on.

The development of a procedure to uniquely identify and store recordings and other physical artifacts (e.g., completed paper-and-pencil instruments) produced during an experimental session will help ensure that no data are misplaced or damaged. One way to do this is to create a session information sheet that can be completed quickly by research assistants after each group session. Provide space on the sheet to record the study name, session date and time, condition, technician's name, comments, and, importantly, a unique code number for each experimental session. This unique code then can be placed on a copy of the file for the session's video recording and all other saved session artifacts. The labeled items, and the record sheet, then can be placed in a large manila envelope or "zip-type" plastic storage bag, and these can then be secured in a locked cabinet or drawer. Such procedures ensure that data are clearly linked and that the confidentiality of the subjects is maintained. Procedures for quickly backing up video recordings can avoid potentially costly disasters. If using digital video, the MP4 or other video file can be transferred by cable to a computer hard drive and burned onto CD or saved on an external hard drive specifically dedicated to the storage of data for the experiment.

Video recording also can be used to record research assistants' interaction with participants. Johnson et al. (2006), for example, recorded each of two experimenters as they ran discussion groups to check that key information in the script was correctly presented and that their demeanor (coded as very confident, mostly confident, mostly hesitant, or very hesitant), tone of voice (coded as firm vs. soft), and use of verbal qualifiers were similar.

## B Using Video Recordings to Deliver Experimental Instructions

Video recordings also can be used to deliver instructions to groups or to single individuals. The advantages of using video recording are increased experimental

control, increased standardization, and ease in running sessions. In a simple setup, participants might sit at study carrels set in rows with a large video monitor placed so that each participant can see it over the top of his or her carrel. A more elaborate setup might have several small testing rooms, each with a video monitor placed where the participant can easily view its screen and a camera positioned to capture a person watching the monitor. In the control room are a computer and, if needed, another camera for delivering “live” instructions to the participants (see later discussion). Cables connect testing room and control room cameras, through a signal switcher, to a computer and monitor. The signal switcher is a piece of hardware that allows signals from multiple sources to be independently routed, quickly, using a simple push-button interface, to any monitor or recorder. The same functions may also be accomplished with computer software if there is a logical break between both live and prerecorded instructions.

Recorded instructions can be presented to participants either as prerecorded or as “live, from another room in the laboratory.” One advantage to presenting prerecorded instructions *as* prerecorded is that it allows for the use of subtitles and graphics, which can emphasize important information. Note that it is important here not to go overboard; use only a few simple effects specifically to stress important pieces of information.

In Hysom’s (2009) research, part of the instructions were described as live because it was necessary for the participant to “introduce” himself to his (prerecorded) partner. In other parts of the session, video was presented explicitly as a prerecorded video “about the abilities we are studying today.” Parts of the video that were presented as prerecorded had much higher production values than did sections presented as live. In acknowledging prerecorded sections, a narrator in a suit and tie sat at a desk in front of a neutral background and spoke directly into the camera.

During key parts of the instructions, subtitles were used to emphasize the material. For example, the words “contrast sensitivity” faded into view along the bottom portion of the screen as the ability was being described. Similarly, when scoring standards were presented, the scene cut from the narrator to a graphic of a scoring chart. As each score category was discussed, the part of the chart containing information about that category was highlighted in a contrasting color. Sections presented as live were recorded in an office setting, with bookcases filled with various equipment and books as the background, and the male researcher wore a lab coat over a white shirt and tie. During experimental sessions, the researcher wore the same clothes, so the transition from in-person interaction (as participants arrived and were taken to their testing rooms) to ostensibly live instructions was seamless.

Introductions between the participant and a prerecorded “partner” were accomplished using a signal switcher. The prerecorded researcher presented as live on the participant’s monitor asked the partner to introduce himself to his partner and then reached off screen and ostensibly pressed a button. The participant immediately saw his “partner” on screen. The partner was prerecorded,

looking into his own camera as he introduced himself by name. The researcher then asked from off screen, “And you are a student here at \_\_\_\_\_? Is that right?” to which the prerecorded partner replied, “Yes, I am a student here at \_\_\_\_\_.” Using the signal switcher, the output from the camera in the participant’s own testing room was immediately routed to the participant’s monitor so that the participant saw himself on his own screen. The researcher’s voice asked the participant what his name was and then verified that he also was a student at the university. Once the participant said his name and identified his university, the screen returned to an image of the researcher in the control room. Because the actor playing the partner can be of any race, age, gender, etc., the use of video introductions allows for the control of status and other characteristics displayed by the partner.

If individuals appearing in a recording are presented as live, they must wear the same clothing for each session. The effort associated with uniformity can be minimized if individuals in the recording are shot only from the waist up. Also, it is important to avoid an outward-facing window in any shot that is ostensibly live because seasons may change.

## C Production of Prerecorded Instructions

When producing prerecorded instructions, it is crucial to prepare thoroughly before beginning to record. As with most aspects of experimentation, the more work that is completed up front, the fewer problems will be encountered in running the study. Previously, we discussed the use and training of confederates. When confederates are recorded, there are some additional considerations. Meet with and train confederates appearing in the video recording to verify, before the scheduled shoot date, that all are fully prepared and can recite or read through their lines without difficulty. Scheduling practice sessions before the scheduled shoot can help make shooting go more smoothly.

During recording sessions, at least one person generally keeps notes about deviations from the script, and multiple shots of each section of the script are made. Minor mistakes (disfluencies, minor deviations from the script) can later be edited out, as needed. (In video presented as live, some disfluencies actually can be expected and may make the recording more believably live.) Once all scenes are recorded, the video recording is carefully reviewed, and the best shots are identified for use in the final instruction recording. When reviewing recordings, the creation of a log, noting the time stamp at which each shot begins, is useful so that later, the experimenter and/or the editor can return quickly to shots chosen for the final recording. Those selected are then edited together so that the final recording flows smoothly.

Video recordings can be produced in house, or all or part may be completed by professionals. The availability of affordable digital video cameras and several commercially available software programs makes possible the in-house production of high-quality video recordings. Most cameras and many of the

currently available software programs are fairly easy to master, at least for basic editing operations. The main advantage of producing recordings in house is that a researcher has more control over the process and, if needed, can make changes to a recording more easily and quickly during pretesting. A potential disadvantage to in-house production is that it requires knowledge of the video software. When production and editing are done by hired consultants from outside the lab, verify how much time they will need to edit and return the recordings. It also will save time and avoid many problems to carefully explain in advance of beginning work exactly what is expected from the consultants and to ask what they require in return. Because experimental instructions are a unique type of video production, additional time should be spent explaining the goals of the recorded communications.

As with live research assistants, the narrator on recorded instructions should appear serious, authoritative, and competent. Distractions should be minimized, and speech should be clear and easy to understand. [Rashotte et al. \(2005\)](#) developed a technique to assess video-recorded instructions, playing them for large classes and asking students for their evaluation of the narrator's presentation using a series of semantic differential items (e.g., too fast/too slow, ignorant/knowledgeable, and competent/incompetent).

Because digital video technology and computer editing programs develop over time, this may affect purchase decisions for equipment and software. Some suggestions here are to standardize as much equipment as possible, including cables, cameras, and media. It may save costs in the long run to avoid proprietary media formats. Note that one useful feature of camcorders is that they can be used as video cameras and as video recorder/players. For cabling, S-video cables and RCA-type audio cables are good choices. One final note about producing video recordings that are to be shown as live is to make sure to turn off the time and date display so that it does not destroy the illusion that a performance is actually live.

## D Using Prerecorded Instructions

Using prerecorded instructions simplifies the running of experimental sessions. It is most useful to have a separate well-labeled digital file for each different condition. Although most of the instructions will be the same for each condition, it may be awkward to switch from one file to another during a session. For each session, after randomly assigning the participant to a condition, the researcher selects the correct file and opens it on the computer desktop or in a location where it will be easy to select.

As with any mechanical or electronic hardware, video recording technology is imperfect. Files get lost or corrupted, computers and monitors break down, and assistants play the wrong recording. The best way to deal with such problems is to avoid them by regularly testing equipment and by having multiple backup files on different forms of media (external hard drive, DVD, etc.) for

each condition. Some errors, if they occur, safely can be ignored. A participant is unlikely to know, for example, if the wrong video file was selected, and if this happens, he or she can most likely simply be moved into the condition that corresponds to the recording that was actually presented. When the worst happens, however, and equipment fails, or some procedural error occurs that is noticeable to the participant, it may be best to stop the session. If such a session must be prematurely ended, debriefing is especially important. In such cases, the participant is informed of what happened, what should have happened, and why it should have happened.

## IX MAINTAINING A SUBJECT POOL

Here, we briefly discuss some of the common problems in maintaining a subject pool, including subject recruitment. For a more detailed discussion of subject recruitment, the reader should consult Kalkhoff, Youngreen, Nath, and Lovaglia in this volume (Chapter 5). For most experiments, research involves undergraduate students attending the academic institutions where the research is conducted. Although experiments also may use older adults, members of the community, or people with institutional attachments such as employees in work organizations or students in schools, we confine our discussion here to subject pools of college students and the unique problems with such pools.

Students are typically recruited from large classes for freshmen. Freshmen and, to a lesser degree, sophomores make the best subjects because they are naive in the ways that are important for experimental research; for instance, they are unlikely to have already participated in other social science experiments. They also should be naive to the hypotheses or goals of the research, and freshmen typically have not had a chance to be exposed to current theoretical work in the social sciences. Subjects also should be naive to experimental design, the possibility of deception, and the manipulation of variables in an artificial setting. Beginning college students also have not had a chance to be exposed to such details of experimental research. In addition, they usually are eager to participate and to do well and are more likely to believe what they are told about the study. On the other hand, using subjects with such a great desire to please the researchers should put experimenters on the lookout for any misinterpretations or possible demand characteristics (discussed previously in the section on deception).

Generally, faculty throughout the university are supportive of academic research, and most are willing to give recruiters access to their students. A letter and an explanation about the study being conducted usually provides the legitimacy needed to make a brief presentation in a class and collect sign-up slips filled out by students expressing an interest in participating in a study. In order not to take up too much of a generous colleague's class time, the recruiting presentation should be brief. The goal is not to educate students about experimental research or about the project. In any recruiting session, researchers are

well advised to give as little information about the project as possible, while establishing their legitimacy as a social science researcher and clarifying the minimal risk for participants in the study.

Experimenters also may want to collect some information on potential participants at the same time. Precious time and subject monies will be wasted if an experiment is run with subjects who previously have been made suspicious by their participation in an experiment or from learning in a class about deception in experiment research. Those subjects can be weeded out by collecting some background information in the recruiting session. Such information also can reveal if the potential subjects are similar in terms of demographic characteristics or if they have had experiences that may put them outside the scope of the theoretical predictions being tested. The trade-off, however, is that the longer the form students are asked to fill out, the fewer the students who are likely to sign up for the project. The researcher's desire for the most appropriate subject pool should balance with the need to get a large (and unselected) subject pool.

An incentive for participation is often required. Some researchers will make an arrangement with colleagues who teach large sections of an introductory course to give extra credit for experiment participation. This may be an effective incentive, but experimenters should be careful not to fill their subject pools with participants from only one major. A better strategy is to recruit from an introductory course or courses that are required of all freshmen. It may be possible to arrange extra credit opportunities with these instructors. Others find that paying subjects will increase the likelihood that they will volunteer to participate and will show up once they have been scheduled.

Pay will need to reflect the going rate at the university or college where recruitment takes place. Not only should it be compared to the minimum wage for working on the campus but also it should be compared to the value students place on other activities, such as studying, working, or socializing. If the value of the other activity outweighs the amount being offered, students may be reluctant to participate and keep appointments. At more affluent universities, if students are more active in sports or organizations or if the pressure for top grades is great, pay will need to be sufficient to lure them away (even briefly) from such activities. Importantly, if rewards are a part of the study (e.g., in many economics studies and in sociological studies of network power), incentives should not be mixed. For example, students who are receiving course credit for a study should not be mixed with students who are being paid because control in incentive structure as an extraneous variable is sacrificed.

## A Scheduling

As discussed in other chapters, some studies may involve a sign-up process for participants and no scheduler will be necessary. However, in studies that require routinized scheduling, a well-trained scheduler with a clear script and a list of what to tell and what not to tell participants can minimize no-shows and reduce

suspicions. Schedulers will need access to long-distance telephoning when scheduling from a student pool because many students likely will provide cell phone numbers with nonlocal (hometown) area codes in their contact information. All subjects should receive two telephone calls—one to arrange the appointment and one, the day of the experiment, to remind of or confirm the appointment. However, before the scheduler telephones anyone, he or she must check the information on the potential participant to ensure that the scope conditions regarding subject characteristics are met. For example, if status-equal groups are being run, females must be grouped with females and males with males.

Schedulers should be provided with a manual similar to the manual for confederates. The manual will tell the scheduler how many and when subjects need to be scheduled. It also will explain how essential it is that subjects know and understand that researchers are counting on them, once they have agreed to participate, at a certain time and that if subjects do not show up, they may jeopardize the entire experiment. This point should be repeated several times to the subjects in the scheduler's script in order to be communicated effectively. (It should not be repeated, however, to the point that subjects feel coerced.)

The scheduler usually includes a very clear introduction identifying himself or herself with the research lab and the particular manner in which subjects were recruited. Subjects then are given a very general description of the study for which they are being scheduled. According to the particular experiment, subjects may be told how long the study takes and the approximate rate of pay.

Subjects should not be scheduled more than a week ahead of the time that they will participate. Things change in subjects' lives in longer time periods, and cancellations and no-shows become more frequent. In the first telephone call, the scheduler should always follow the script. If the potential participant is unable to participate during a time slot on the schedule, the scheduler will have clear instructions on what to tell him or her. The scheduler may say that he or she will call back another week and try again to schedule, or the scheduler may say that he or she will telephone again during the next semester or session. This alerts potential participants to the possibility of another telephone call and, if they do not object to the scheduler telephoning them again at a later date, the scheduler will have a much better chance of keeping them in the subject pool.

Schedulers also will need to give clear directions to the building and room where subjects are to meet experimenters. It is best to repeat this information several times during the initial telephone call, try to get the subject to repeat it back, and repeat it in the follow-up reminder call. Subjects generally will not write down the information, so repetition is the best way to get them to remember the directions. Another way to help subjects to remember appointments is to create an opportunity for them to write it down. One scheduler's script that we use states, "In a moment I'll give you some important information. While I place you on our schedule, can you get a pencil or pen and paper so that you can take down this information?" After the participant returns, the appointment time is confirmed and directions are given.

The directions also should include special instructions for entering the building (if this is in the evening or on a weekend and doors are locked) or where to meet the experimenter if the participant cannot find the room. Subjects also can be given a telephone number at which to reach an experimenter if they must cancel their appointment. Sometimes subjects do write this number down or remember the name of their contact person. With the callback feature on cell phones and most land-line telephones, however, schedulers also find themselves being contacted at the number from which they originally scheduled the experiment appointment.

Schedulers will need to keep careful lists of who is scheduled for which time. The day a session is scheduled, the scheduler should make a second telephone call or send a text message to remind subjects about their appointments. This reminder call is especially important because potential subjects often will forget that they have made an appointment to participate. With cell phones, appointments may be made while participants are unable to write them down. A reminder call helps reduce the proportion of no-shows. The follow-up reminder phone call may be left on voice mail. The message need only include the time of the appointment and where subjects are to go along with a telephone number for cancellations. It is essential that everyone who is scheduled be contacted.

It is also important to mention that the experiment may begin with the appointment call made by the scheduler. In the status-construction experiment, for example, potential subjects were told that they would be working with a partner so it was essential that they show up on time in order not to keep anyone waiting.

## B No-Shows

“No-shows” is the term for subjects who fail to make the experiment appointment and do not call to cancel. The experimenter must decide if he or she will try to recontact the potential participant and reschedule the experiment appointment. In general, unless our subject pool is dangerously small and we may run out of potential subjects, we do not recontact no-shows, but this is a matter for the experimenter to consider. Although there are legitimate reasons for not making a scheduled appointment, we reason that a no-show may be less responsible and/or too busy to participate, and that the likelihood that he or she will make the effort to keep the second appointment is not great.

Student culture in terms of socializing activities will have an effect on the likelihood of no-shows and will need to be considered when schedules are set up. In general, no-shows are most likely to occur on Friday afternoons (or even Thursday afternoons) when students are beginning their weekend relaxation and partying. They also are most likely to occur at the midpoint and end of the semester or term when students are preparing for midterms or finals and writing papers and easily may overestimate the amount of free time that they will have. Even weather can affect the no-show rate. Rainy days and days with beautiful

weather are worst; cloudy days seem to be best. In addition to using reminder telephone calls, the number of no-shows also can be reduced by a skilled scheduler who keeps track of experiment times that are consistently difficult to fill or that potential subjects are reluctant to commit to. Some modifications of the experiment schedule may be needed to reduce no-shows and late arrivals.

What about overbooking, which airlines and restaurants do routinely? We have used this technique, although it can deplete a subject pool quickly. If an experimental design requires several subjects for a group, such as a six-person discussion group, then overbooking may be the only way to keep the project running. If exactly the right number keep their appointments, of course all is well. If an “extra” person arrives, we greet that person, explain that because of (unstated) circumstances we need to reschedule him or her, and express apology for inconveniencing him or her. We also pay a small amount for showing up and reschedule the person right then (to be sure that he or she will get into the new group). We have not found any resentment at coming once without participating, and these people are very likely to keep the second appointment and to be fine subjects. Thus, the only negative aspect of overbooking is that it uses a subject pool faster. It does not seem to cause harm to individuals who come more than once or to affect their behavior in the experiment when they do participate.

## C Contamination of the Pool

The subject pool may become contaminated if potential subjects receive biasing information beforehand. One non-obvious source of that problem is faculty who teach the class from which subjects are recruited. It is important that faculty do not give subjects negative information that would make them reluctant to participate. It is also important that faculty do not accidentally slip and tell subjects what the experiment is really about or, if there is deception, what it might involve. For this reason, it is best that as few people as possible have such information about the study. Colleagues should know the subject of the investigation only in the most general terms. Colleagues also will have respect for experimental research when recruiters are respectful of faculty time in the classroom and make their presentations brief. It is particularly damaging to recruitment rates if the instructor introduces recruiters as people “who are looking for guinea pigs.” Try to get the instructor to allow you to introduce yourself, or at least ask him or her to simply give your name and say you are looking for help with an interesting research project.

The subject pool also may become contaminated if subjects who have participated in the study tell potential subjects what the hypotheses are, that deception is involved, or that they will work with a confederate. The experiment should be a pleasant experience for subjects so that they will be sympathetic to the goals of the research once they have participated. They should be treated with respect, paid for their work, and not kept beyond the time indicated. If the experiment must be canceled for any reason (equipment failure is the most

common), it is worth the money to pay subjects for some of their time in order to generate good will and maintain a professional reputation on campus.

Some experimenters have found that subjects are less likely to reveal manipulations or deception if they are made to feel a part of the study. In the postsession debriefing, not only are subjects protected by the experimenters revealing the purpose of the research, but also subjects are included in the research. Experimenters may include them by asking if subjects understand the predictions and if subjects think that the predictions will be supported. If subjects are allowed to comment on and share personal experiences about the variables under investigation, they are more invested in the success of the study and are more likely to keep it secret. To minimize this problem, we include the following statement in the experimenter script for our postexperimental interviews:

*Also, we would appreciate it if you would not tell your friends about what we are looking for in this study. Once people know the hypotheses, they don't behave the way they normally would in an experiment. Since we are recruiting from freshman lecture classes, you will probably know others who will be participating in the study. If they ask, please only tell them that it's a study about perceptions in groups and nothing more. It's fine to discuss the study in more detail once they've finished participating.*

The pool should be repeatedly assessed by experimenters for possible contamination, and potential problems should be anticipated. If the size of the pool reaches such a low level that random assignment to conditions becomes difficult, it may be possible to contact additional faculty members and recruit from additional classes, increasing pool size. However, experimenters should be careful that all students at a particular level are included in the pool initially and that all have an equal chance of being selected. This helps to fulfill the random assignment requirements as well as meet ethical concerns about volunteers being able to participate.

## X PAYMENT AND CREDIT ISSUES

As noted previously, participants are typically paid a small amount of money to encourage their participation. The establishment of a “petty cash” fund, maintained in the lab, from which to pay participants is common practice. Depending on the amount paid to participants and the total number of participants in a study, the amount of cash needed for payments may be substantial. Financial services personnel may not be familiar with social science experimentation and may not, therefore, understand why a petty cash fund of such a large size is necessary. In one case we know of, a researcher was asked to submit information for each subject, requesting for each that a separate check be issued and mailed to participants. Considering the modest sums typically paid to participants in this type of research, that procedure, in addition to being costly, would be likely to make the payment less of an incentive. With this in mind, it may be useful to

get to know your financial services personnel, the department funds custodian, and any other individuals involved in the process well before requesting funds. Positive interpersonal relations can be crucial in quickly resolving problems or confusion about funds if it arises.

Different institutions and funding agencies may have different rules and procedures for establishing and maintaining a petty cash fund. The first time you establish such a fund, it may take longer and be more complicated than you anticipate. Typically, procedures require a named “custodian” and the use of prenumbered receipts. The custodian is most often the faculty member conducting the study. Funds are released to the custodian (with options for payment to be made in cash or check or electronically), who has legal responsibility for the funds. After spending part of the money, the custodian then presents valid receipts and/or cash, usually by a date set at disbursement, equal to the amount initially released. Policies also usually stipulate that funds be kept in a secure location, where access is limited to laboratory personnel; a locked file drawer in a locked laboratory room meets that requirement.

Research assistants will need to be trained to issue numbered receipts in the correct order, verify that participants correctly complete their part of the receipt, return the cash box to its secure location after each session, and lock the lab. Research assistants who are running the study will need to keep track of the funds so that funds can be replenished before they run out. When petty cash needs to be replenished, receipts totaling the amount initially disbursed and any required paperwork are returned to the institution. Turnaround time for replenishment may vary, but 2 weeks is common. If someone forgets to replenish petty cash in time, an experimenter may need to be prepared to stop running subjects for a few days.

To avoid delays and other problems, it is important to verify that all expenses are entered into proper categories. With this in mind, determine what expenses are allowable, and note any rules that differ between funding agency and the university. A researcher has to incur only expenses that are allowed by both sets of rules. A common rule is not to commingle petty cash with other funds. For instance, nobody associated with a project should regard the petty cash box as a place to borrow money. Also, multiple accounts, such as a research petty cash fund and a departmental petty cash fund, should not be mixed. For questions about allowable uses of funds, check with the funding agency directly—preferably with the officer in charge of the grant, if there are questions regarding a given expense. (Remember that such officials are usually busy people in charge of many grants, and so the grant number should always be included in any communications. Also try not to refer every question to the program officer. Handle what you can on your own, reserving contacts over handling money for cases that really cannot be dealt with locally.)

In the case of petty cash or working funds, it is likely that a fund’s allowable uses are stipulated when the fund is established, and other uses will not be honored. It will be helpful to know the project (account and fund) numbers

associated with the experiment. In larger departments, this information is most likely available from the department accountant; in smaller departments, the office manager may have this information. Finally, request amounts that are multiples of the amount you will be paying each participant, and get the correct denominations initially so that research assistants do not have to scramble for change during experimental sessions.

## XI EXPERIMENTER EFFECTS

Experimenter effects are errors introduced during the collection or analysis of experimental data due to the behavior of the experimenter. They can affect the data collected in an experiment and thereby confound the analysis of results in at least four ways (*Rosenthal, 1976*) through: (1) subtle differences in participant treatment; (2) errors in recording data; (3) errors in selecting cases; and (4) errors in the analysis of data. In the vast majority of cases in which experimenter effects have been studied, these effects bias results in favor of the hypothesized results, with the experimenter unaware that he or she is even making an error or treating participants differently based on expectations.

*Rosenthal (1976, 2005)* suggests several strategies for addressing experimenter effects. The best known is the use of experimenters who are blind to the condition being run. We mentioned using blind interviewers previously with reference to assessing the extent to which scope conditions are not instantiated for a given subject. When an experimenter is blind to the condition, his or her expectations regarding the hypothesized outcome for the condition are controlled, and so too is any cuing behavior caused by such expectations. The careful development and assessment of standardized prerecorded instructions can have the same effect of controlling for differences in experimenter behavior since the behavior of the experimenter, shy of the manipulation of the independent variable, is the same for every participant.

Another technique for the control of experimenter effects is to minimize contact between the experimenter and the participant. The less exposure a subject has to experimenters, the less likely it is that cues from the experimenter will be “picked up” by the respondent. Prerecorded instructions greatly reduce the amount of contact between participants and experimenters. Because any errors or problems in a video recording will affect all participants, however, careful pretesting and development of these instructions to identify and rectify problems is that much more essential.

## XII CONCLUSION

Conducting experimental research to test theories or construct theoretical explanations requires careful attention to all details from setting up the lab to developing procedures to implementing the experiment. Published summary reports of experimental tests may lead to conclusions that the data collection process

was relatively easy because experimental research is often described in a clear and simple presentation. Readers of this chapter, however, likely will conclude that setting up an experiment, training assistants, and recruiting subjects are far more complicated and time-consuming than a research report indicates. Underlying experimental operations are a variety of hidden problems and issues. In this chapter, we suggest solutions that we and others have found helpful for some of the common problems and issues. Through sharing the details of our experiments, we hope not only to facilitate replication of social scientific research but also to encourage laboratory experimentation for building social scientific knowledge.

## REFERENCES

- American Sociological Association. (1999). *Code of ethics and policies and procedures of the ASA committee on professional ethics*. Washington, DC: American Sociological Association.
- Berger, J. (2007). The standardized experimental situation in expectation states research: Notes on history, uses, and special features. In M. Webster, & J. Sell (Eds.), *Laboratory experiments in the social sciences* (pp. 353–378). Burlington, MA: Elsevier.
- Berger, J., Fisek, M. H., Norman, R. Z., & Zelditch, M., Jr. (1977). *Status characteristics and social interaction: An expectation-states approach*. New York: Elsevier.
- Berger, J., Zelditch, M., Jr. (1977). Status characteristics and social interaction: The status-organizing process. In J. M. Berger, M. H. Fisek, R. Z. Norman, & M. Zelditch Jr., (Eds.), *Status characteristics and social interaction: An expectation states approach*. New York: Elsevier.
- Cook, K., Cronkite, R., & Wagner, D. (1974). *Laboratory for social research manual for experimenters in expectation states theory*. Stanford, CA: Stanford University Laboratory for Social Research.
- Dowrick, P. W., & Biggs, S. J. (1983). *Using video: Its psychological and social implications*. Chichester, UK: Wiley.
- Hegtvedt, K. A. (2007). Ethics and experiments. In M. Webster & J. Sell (Eds.), *Laboratory experiments in the social sciences* (pp. 141–172). Burlington, MA: Elsevier.
- Hysom, S. J. (2004). *An experimental test of the theory of reward expectations*. PhD dissertation, Atlanta: Department of Sociology, Emory University.
- Hysom, S. J. (2009). Status valued goal objects and performance expectations. *Social Forces*, 87, 1623–1648.
- Johnson, C. (1994). Gender, legitimate authority, and leader–subordinate conversations. *American Sociological Review*, 59, 122–135.
- Johnson, C., Fasula, A. M., Hysom, S. J., & Khanna, N. (2006). Legitimacy, organizational sex composition, and female leadership. *Advances in Group Processes*, 23, 117–147.
- Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, 17, 776–783.
- Orne, M. T. (1969). Demand characteristics and the concept of quasi-controls. In R. Rosenthal, & R. L. Rosnow (Eds.), *Artifact in behavioral research* (pp. 147–179). New York: Academic Press.
- Rashotte, L. S., Webster, M., Jr., & Whitmeyer, J. M. (2005). Pretesting experimental instructions. *Sociological Methodology*, 35, 163–187.
- Ridgeway, C. L., Boyle, E. H., Kuipers, K. J., & Robinson, D. T. (1998). How do status beliefs develop? The role of resources and interactional experience. *American Sociological Review*, 63, 331–350.

- Roethlisberger, F. J., & Dickson, W. J. (1939). *Management and the worker*. Cambridge, MA: Harvard University Press.
- Rosenthal, R. (1967). Covert communication in the psychological experiment. *Psychological Bulletin*, 67, 356–367.
- Rosenthal, R. (1969). Interpersonal expectations: Effects of the experimenter's hypothesis. In R. Rosenthal, & R. L. Rosnow (Eds.), *Artifact in behavioral research* (pp. 187–227). New York: Academic Press.
- Rosenthal, R. (1976). *Experimenter effects in behavioral research*. New York: Irvington.
- Rosenthal, R. (2005). Experimenter effects. In K. Kempf-Leonard (Ed.), *Encyclopedia of social measurement; Vol. 1*. (pp. 761–875). New York: Elsevier.
- Sell, J. (2008). Introduction to deception debate. *Social Psychology Quarterly*, 71, 213–214.
- Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behavior. *European Journal of Sociology and Psychology*, 1, 149–177.
- Troyer, L. (2001). Effects of protocol differences on the study of status and social influence. *Current Research in Social Psychology*, 16, 182–204.
- Troyer, L. (2007). Technological issues related to experiments. In M. Webster, & J. Sell (Eds.), *Laboratory experiments in the social sciences* (pp. 173–191). Burlington, MA: Elsevier.
- Walker, H. A., & Cohen, B. P. (1985). Scope statements: Imperatives for evaluating theory. *American Sociological Review*, 50, 288–301.

## Part II

# Experiments across the Social Sciences

Part II consists of 12 chapters that span several social science disciplines. Although experimental research entails some of the same issues in all disciplines (e.g., those described in Part I), other issues are more prominent in the types of experiments conducted in particular disciplines. Represented here are sociology (Chapters 8, 9, 11, and 12), economics (Chapters 15 and 16), political science (Chapters 13 and 14), psychology (Chapter 18), organizational behavior in business (Chapter 19), and communication (Chapter 17). Social dilemmas, in Chapter 10, have been studied in several disciplines, and this chapter discusses similarities and differences and integrates materials across several disciplines. All of the chapters in Part II ask and answer ways that experimental designs are useful for the particular kinds of issues in several social sciences.

Chapter 8, by Morris Zelditch, Jr., identifies different eras in sociological experiments, defined by the goals of practitioners. Early experiments were “effect experiments”—studies whose main aim was to show that some process, such as conformity or norm creation, could be created in a laboratory. The growth of sociological theory in the past approximately 50 years led to theory-testing experiments, and theoretical research programs spawned the modern sequential, theory-driven experimental designs of the kind described in the chapters of this book. Zelditch shows how theory-driven experiments deal with issues of external validity, making experiments a good source of knowledge for situations far beyond the laboratory.

Chapter 9, by Linda D. Molm, describes research using social exchange theories from their outset through the present, including her own research designs and approach over several decades. This field developed from demonstrating that reward–cost analyses could be applied to many social interactions to explicit theoretical descriptions for different forms of exchange and the formation of associated effects such as trust and reciprocity. Early experiments focused on two-person exchanges but quickly

moved to studies of alliance formation in social networks of differing sizes and connectivity. Molm describes several basic research designs—for negotiated, reciprocal, generalized, and productive exchanges—and shows how they can be modified to investigate a wide range of phenomena involving social exchange.

Chapter 10, by Jane Sell and Bruce Reese, treats social dilemma experiments as studied from across social science disciplines with emphasis on political science, economics, sociology, and psychology. These experiments study one of the oldest problems in social science—namely how to provide benefits available to everyone supported by voluntary contributions (e.g., public radio) while at the same time minimizing the problem of “free-riders” who use the benefit without contributing to it. Sell and Reese describe different experimental designs that seek to isolate different explanations for factors affecting cooperation, describe some of the most important results, and identify promising areas for further investigation.

Chapter 11, by Martha Foschi, offers a close examination of terms involved in experimentation that can be misunderstood. She considers the parts of a hypothesis and how those relate to empirical research, which she illustrates with examples from her own research program over several decades on double standards. Foschi considers what operationalization means in the context of theory-testing experiments, making the important distinction of operationalizing theoretical variables from operationalizing accidental features of an experiment. She offers a set of design features that require manipulation checks, and she discusses how to assess those checks.

Chapter 12, by Joseph Berger, describes his and his colleagues’ work of more than half a century creating and developing an experimental situation used by large numbers of social scientists worldwide to study status and expectation processes. In addition to providing a fascinating narrative in the history of social science, this chapter identifies design considerations and methods that are relevant to any experimental design.

Chapter 13, by Rose McDermott, describes a well-known large-scale study of effects of hormone replacement therapy on women’s health, identifying design errors that made results difficult to interpret. She then traces experimental studies in political science dating back almost a century in the United States, focusing on the difficult problem of attributing cause in the highly complex situations of interest to political scientists. Early experiments were guided by a desire to reduce the number of statistical interactions so that researchers could study a few factors at a time. Recent and contemporary experiments have generated considerable information on some phenomena, such as voting, and political scientists have developed ingenious ways to study sensitive topics and to relate field experiments to laboratory experiments. She concludes by assessing contemporary studies

in political science, including new areas for investigation and new technology useful for experiments in political science.

Chapter 14, by Rick K. Wilson, describes how voting and agenda setting have been investigated experimentally within political science. The “canonical experiment” on voting in political science is one in which a group of individuals with disparate interests vote repeatedly until they reach agreement (or give up). Voting experiments illustrate both kinds of outcomes, equilibrium (in which members reach an agreed-upon outcome) and nonequilibrium (in which they do not). Agenda-setting experiments consider how groups decide to allow votes on particular issues. Wilson notes that experimental research in voting behavior shows the importance of sociological and psychological factors along with the political. In this way, experimental political science may have similarities to experimental economics in demonstrating how theories may benefit from incorporating processes previously studied in other disciplines.

Chapter 15, by Catherine Eckel, examines the recent history of experimental economics. Experiments are relatively recent in economics and were initially met with skepticism because often results did not match predictions from well-accepted theories. As Eckel describes, however, anomalies discovered in experiments played important roles in theoretical development for game theory and expected utility theory. In some cases, anomalies led to entirely new lines of theory growth. Development of experiments in economics was fostered by certain individuals, whose contributions Eckel recognizes here. She examines canonical games in economics—the double auction, public goods, ultimatum, and trust—and shows how economic understanding has advanced through their use. The chapter concludes by identifying some areas of wide interest and integrating the history with an assessment for future developments of experimental research in economics.

Chapter 16, by Giovanna Devetag and Andreas Ortmann, describes a widely used subset of experimental economics experiments: coordination. They outline the four general types of coordination games—pure coordination, Pareto ranking, mixed motive, and critical mass—and show how they have been used to answer a wide range of theoretical questions. All address the general problem fundamental to social action: how can organization be accomplished? Devetag and Ortmann analyze and summarize how experimental research has advanced the theoretical field, and they identify promising topics for future investigation.

Chapter 17, by Srividya Ramasubramanian and Chantry J. Murphy, describes and analyzes experimental studies of media stereotyping. Within media, critical concepts relate to the audience or audiences involved and the media transmission itself. These authors detail experimental techniques

designed to answer questions related to how stereotyping occurs through media presentation. They note that different types of media have differing effects, in part because interaction between the message and the audience varies: an interactive game is a different context than a newspaper or a television show. They conclude by noting new arenas for investigation, especially in further analysis of the processes of stereotyping, rather than only its effects.

Chapter 18, by Michael K. Lindell, examines the judgment and decision-making literature, developed primarily in psychology and economics but then utilized in a wide variety of disciplines and approaches. Lindell details how this literature has evolved through extensive use of ingenious experimental designs to systematically vary factors such as alternatives to the decisions, familiarity with decision-making, and the media by which information is provided. As he notes, most of this literature evolved from the concept of rational decision-making and utility models. Although such models were powerful, they were often not adequate descriptors of how individuals actually made decisions. He concludes by suggesting fruitful areas for new investigations and expansions.

Chapter 19, by Stefan Thau, Marko Pitesa, and Madan Pillutla, shows how theoretical issues in business, especially understanding organizational behavior, can be advanced through the use of experiments. They note that many organizational problems involve issues of interdependence. Examples include coordination failures (also addressed in Chapter 16), lack of trust (also addressed in Chapter 15), and free-riding (also addressed in Chapter 10). Although most organizational behavior research involves observational studies rather than experiments, these authors argue that experiments can eliminate alternative hypotheses and advance applied research. They focus on a particularly important problem, one that is sometimes difficult to assess empirically: unethical behavior in organizations. They describe field and experimental studies of processes and structures involved in producing ethical and unethical behavior.

## Chapter 8

# Laboratory Experiments in Sociology

Morris Zelditch, Jr.

*Stanford University, Stanford, California*

## I INTRODUCTION

The 1950s saw rapid and prolific development of theory and research on “small groups.” The 1980s and 1990s saw even more rapid, more prolific development of theory and research on “group processes.” The difference in how the field was described is trivia, perhaps, but in some ways an instructive comment on a sea change in the nature and function of experiments between the two periods. The 1950s were in fact a watershed in the history of experiments in sociology, marked by a considerable climb up the ladder of abstraction and, with it, a considerable change in the nature of the programs that grew out of them—a change that, in turn, meant not only more programs but also more growth in them.

## II EFFECT EXPERIMENTS

The Asch experiment was one of the classic experiments of the 1950s. In the Asch experiment, a naive subject (S) made a sequence of choices between one unambiguously correct and two unambiguously incorrect stimuli in the face of unanimous and incorrect responses by seven peers, all of whom were confederates of the experimenter (E). The experiment consisted of a sequence of trials, each of which required S to match a standard with a line, among three comparison lines, of the same length as the standard. S was instructed to respond orally, and the oral response of six of the seven confederates preceded S’s response. The instructions led S to believe that he or she was participating in an experiment on visual perception, so there was motivation to respond correctly. On the other hand, the unanimity of the confederates exerted pressure on S to conform to an incorrect response. Individuals in a control group in which there were no confederates made their choices with almost complete accuracy. A unanimously incorrect majority deflected a third of S’s choices in its direction, and 75% of Ss made at least one error in the presence of the majority (Asch, 1951).

Contemporary experiments in sociology mostly test, refine, or extend a theory or test its application. The Asch experiment did not so much test a theory as demonstrate an effect. The effect it demonstrated was important and powerful. However, it was also complex and underanalyzed: was it due to group pressures for conformity, which would imply that Ss knew they were making incorrect responses but went along because they were sensitive to the attitudes of the others in the group? Or was it due to cumulative evidence that they could not trust their own senses and therefore had to trust the senses of others to make a correct choice, which would imply that Ss believed they were making correct responses (Deutsch & Gerard, 1955)? The experiment demonstrated an effect but did nothing to sort out its causes.

Many other classics, either published in or still read in the 1950s, were effect experiments. For example, in Lewin's group decision experiment (Lewin, 1947), a persuasive communication, in either a lecture or a group discussion, attempted to change customary habits such as eating practices. The persuasive communication concluded with a request for a decision, made publicly, a specific period to act on it, and information that a follow-up would be made, after which the experiment's effect was measured. Group discussion was more effective than a lecture in changing habits. However, was the effect due to group discussion, making a decision, hence a commitment, making the commitment public, or the fact that the decisions made in the course of group discussion were consensual (Bennett, 1955)?

It may be difficult, even contentious, to construct canons, but surely Sherif's (1935) autokinetic experiment and Bavelas's (1950) communication networks were also among the signature experiments of the period. Surely, if there was a canon, it included Bales' laboratory observations of evolving group structures (Bales, 1953; Bales & Slater, 1955; Bales, Strodtbeck, Mills, & Rosebourough, 1951). All of these, like the Asch experiment, demonstrated important and powerful but complex, underanalyzed effects.

### III EFFECT RESEARCH PROGRAMS

A single experiment is neither precise enough nor rich enough to leave no questions unanswered. Answering its unanswered questions gives rise to a research program—a series, typically a sequence, of interrelated experiments. Because the effect it demonstrates is complex and underanalyzed, the research program to which an effect experiment gives rise is often concerned, first of all, with explicating it—that is, with analyzing what exactly it was that had the effect.

“Complex” means that more than one process was involved in one concrete effect. The idea of “a” process may be philosophically difficult, but, generally speaking, it means regularity of an effect—that, given the same conditions, the same causes have the same effects. “Underanalyzed” means that conceptualization of the effect has not distinguished two or more processes from each other.

Hence, experiments demonstrating underanalyzed, complex effects confound them—the same causes, under the same conditions, therefore have different effects.

Whether an effect *is* in fact complex is an empirical question, open to experiment. Thus, the Asch experiment led to experiments such as [Deutsch and Gerard's \(1955\)](#), which argued that Asch confounded the normative effect of group pressures for conformity with the informational effect of others as sense evidence. Distinguishing the two mattered because the former but not the latter depended on the existence of a group, whereas the latter but not the former would be found whether judgments were anonymous or not. These differences were confirmed by replicating the experiment in the same setting but redesigning it, distinguishing the effect of a group versus an aggregate; of public versus private, anonymous judgments; and of privately versus publicly committing oneself to them.

In the same way, Lewin's group decision experiment was followed by experiments such as [Bennett's \(1955\)](#), which replicated but redesigned it to distinguish the effect of group discussion from the effects of having to make a decision, having to make it publicly, and consensus. Results showed that even without group discussion, commitment and consensus were capable of generating the effect Lewin had found.<sup>1</sup>

However, effect programs are also concerned with the causes, conditions, and mechanisms of the effect. Because the effect is complex and underanalyzed, the causes, conditions, and mechanisms they investigate are frequently ad hoc and unconnected, the search for them open-ended, the outcome incoherent. For example, an experiment by [Emerson \(1964; see his “first experiment”\)](#) sought an explanation of the Asch effect in three unrelated causes, two of them more or less ad hoc: in the motivation to participate in the group, derived from Festinger's theory of informal pressures toward uniformity ([Festinger, 1950](#)); in group expectations, derived from “commonplace” common sense; and in status insecurity, more or less on a hunch. The incoherence of the causes, conditions, and mechanisms sought is simply a by-product of the effect's complexity. In consequence, although the growth of the Asch program has been voluminous, it has, like the effect, been an incoherent mix of unconnected causes, conditions, and mechanisms ([Allen, 1965, 1977](#)).<sup>2</sup>

## IV THEORETICALLY ORIENTED EXPERIMENTS

Not all of the experiments of the 1950s were effect experiments. A few tested theories, even a few “classics” such as [Back \(1951\)](#) and [Schachter \(1951\)](#), both of which tested [Festinger's \(1950\)](#) theory of social pressures toward uniformity

1. Among many other examples, see also [Burke \(1967\)](#) and [Lewis \(1972\)](#), following Bales; [Mulder \(1960\)](#) and [Faucheux and McKenzie \(1966\)](#), following Bavelas.

2. For another example, see [Glanzer and Glaser \(1961\)](#) on the evolution of Bavelas' program.

in informal groups. The 1950s were in fact a time of change—a time during which experiments were increasingly becoming oriented to testing, refining, and extending theories or testing their applications. Some of these experiments emerged out of the explication of an effect, as Berger and Conner (1969) did from Bales. Their experiment was a test of a theory, not a demonstration of an effect. However, the theory it tested was an analytic deconstruction of the effect. It took an effect demonstrated by Bales (Bales, 1953; Bales et al., 1951; Bales & Slater, 1955) as its starting point, but it was both more abstract and simpler than the effect Bales had demonstrated. Bales found that initially undifferentiated groups evolved inequalities in rates of participation; who spoke to whom; who liked whom; who was asked for orientation, opinions, and suggestions; who offered them; who agreed with them; and who made overtures to others, expressed antagonism, showed tension, or released it. For the most part, these inequalities were highly intercorrelated and, once emerged, tended to be stable. Berger (1958) conceptualized them more abstractly as action opportunities, performance outputs, unit evaluations, and influence. Because they were highly intercorrelated, they were regarded as one hierarchy of power and prestige. No attempt was made to encompass Bales' "social-emotional" categories because they were not highly correlated with the power-prestige order, and no attempt was made to explain them.<sup>3</sup> Analytical simplification of Bales' effect, the purpose of which was to isolate a process, simply left out the elements of other processes.

About this more abstract, analytically simplified process, Berger (1958) reasoned that the emergence of a power-prestige order reflected an underlying structure of expectations for performance of a collective task. Differentiated expectations emerged out of differential evaluations of performance when there were disagreements because disagreements had to be resolved in order to reach a "group" decision. Once emerged, these expectations probabilistically determined the observed power and prestige order, the elements of which were highly intercorrelated because they were all functions of the same underlying expectation states. Because the expectations both determined and were determined by the observed power and prestige order, any change in it was itself a function of the order; hence, it was likely to maintain itself unless and until disturbed by some change in the conditions of the process.

Experiments that test theories are, like the theories they test, analytic simplifications of a more concrete phenomenon. Thus, Berger and Conner's (1969) work is essentially an analytic simplification not only of Bales' effect but also of his methods of observing it. For example, Bales observed open interaction; phase 2 of Berger and Conner's experiment controlled it. Their experiment had

---

3. The fact that they did not correlate with his task categories had led Bales to infer that role differentiation was a part of the process. Later, Lewis (1972) found that "role differentiation" was true only of homogeneous groups, and Burke (1967) found that it had to do with the legitimacy problems of emergent hierarchies.

two phases, in the first of which pairs of university students were publicly given scores on a test of a fictitious ability. The test consisted of repeated trials presenting sets of three words, one in English and two in a fictitious language. Ss were told that one of the two non-English words had the same meaning as the word in English and that, by comparing the sounds of the non-English words, they would be able to decide which meant the same as the English word. The ability to do this was called “meaning insight ability.” The scores, as fictitious as the test and rigged by E, were interpreted to them as either “exceptionally superior” or “exceptionally poor.”

Because feedback of the scores in phase 1 was public, S knew both self and other’s scores. This created four experimental conditions: (1) an S whose meaning insight ability was exceptionally superior but whose partner’s was exceptionally poor, or, more simply, an S whose performance-expectation state was high self–low other; (2) high self–high other; (3) low self–low other; and (4) low self–high other. The task in phase 2 also consisted of repeated trials, each of which presented sets of three words, except that only one was in the same fictitious language as in phase 1 and two were in English. The task was to decide which of the two English words had the same meaning as the non-English word. In phase 2, however, selection of a correct answer required three stages. Every time Ss were presented with a set of alternatives, Ss first made a preliminary selection, exchanged their initial choices with their partner, and then made a final choice. The Ss could not verbally communicate nor even see each other but indicated their choices to E and to each other using a system of lights and push-button switches. Except for 3 of a total of 25 trials, Ss were led to believe that their initial choices disagreed. The purpose of exchanging information was defined as determining how well they worked together as a team. The final choice was completely private. Ss were told, moreover, that it would be evaluated in terms only of a team score, which was simply the sum of the number of correct final choices each made, and would not record, hence not reveal, their relative contributions to the score. Because the choices were binary, the final choice in the 22 disagreement trials indicated either acceptance of or resistance to the influence of the other. The result was that the probability of acceptance of the influence of the other was greatest in the low self–high other condition, least in the high self–low other condition, and approximately equal in the high–high and low–low conditions.

This experiment exemplifies an increasing number of “contemporary” experiments that increasingly test a theory and, like the theory they test, analytically simplify a phenomenon by focusing on only one of its processes. However, like this particular example, any theory-oriented experiment tests only one, or only a few, of the implications of a larger theoretical structure that has other implications. A theoretically oriented experiment is no more likely than an effect experiment to leave no unanswered questions. However, the questions it leaves unanswered are questions about a theory, not an effect. This makes a significant difference in the kind of research programs that emerge out of the answers to

them. The increasingly theoretical orientation of contemporary experiments has therefore led to an increasingly theoretical orientation of contemporary research programs.

## V THEORETICAL RESEARCH PROGRAMS

Answering a question left unanswered by a theoretically oriented experiment modifies a theory, not an effect. The outcome is in fact a theory. For example, a question left unanswered by Berger and Conner (1969) was how the process would behave if a group were initially differentiated, for example, by race, gender, education, or occupation. Answering it led to the theory status characteristics (Berger, Cohen, & Zelditch, 1966). In the standard model of theory growth, the new theory subsumes, hence displaces, the old theory—an account in which theory growth climbs a neat, simple, linear path. However, that is not always what actually happens. There are a number of different ways one theory can be related to another (Berger & Zelditch, 1997), and whether new theory displaces old theory—and therefore how a program grows—depends on how they are related. Status characteristics theory did not displace the power–prestige theory. The difference in the conditions under which the process occurred led to a difference in its effect, accounting for which led to an auxiliary theory—of status generalization—not found in the power–prestige theory. The two accounted for different effects under different conditions. On the other hand, the theory of status characteristics shared much in common with the theory of power and prestige—for example, the same basic ideas about expectation states. Both were expectation states theories, part of the same program. Thus, what had been *a* theory became a family of theories. The power–prestige theory simply proliferated, differentiating into two distinct but related theories.

Some theories, on the other hand, *do* displace earlier theories. For example, the initial theory of status characteristics was a theory of the effect of a single characteristic (Berger et al., 1966). Extension to multiple characteristics (Berger, Fisek, Norman, & Zelditch, 1977) displaced it. That is, the multicharacteristic theory was capable of explaining anything explained by it and more; hence, it was superfluous. Thus, in the course of its growth, the program gradually came to be made up not just of two theories but two continually evolving branches, each elaborating a different but related theory. Other branches emerged as other unanswered questions were asked, such as the following: What was the effect on the formation of expectation states of legitimate sources of evaluations, such as teachers (“source” theory)? Or of expectations held by other interactants (second-order expectation states)? What was the effect on both the formation of expectations and the power–prestige order of relations between performance and reward expectations (the theory of distributive justice)? Or the effects of nonverbal interaction (e.g., status cues) on the formation of expectation states? Or of legitimation on the power–prestige order? New branches also emerged out of integrating theories in already existing branches,

such as the status characteristics and source theories. Finally, they also emerged out of applications of the theory—for example, of status characteristics theory and source theory to schools and also status characteristics theory to gender.<sup>4</sup>

This is obviously a complex structure, and I have not even touched on the theoretical and methodological strategies that guided the construction and application of its theories. Not only is it complex but also other examples often look very different because different theory–theory relations give rise to different patterns of growth and hence to programs with quite different structures. What they all seem to have in common is that they are all made up of a set of theoretical and methodological strategies, a network of interrelated theories embodying them, and empirical models interpreting its theories, together with a body of theoretical and applied research testing, refining, and extending the program’s theories and their applications ([Berger & Zelditch, 1997](#)).

Although I have illustrated them by describing only one example, expectation states theory is merely one of an increasing number of theoretical research programs in sociology. I have not made a census of them, but I have been teaching them, providing students with examples of them, and coediting anthologies of them for most of the period I am describing, and there are considerably more of them now than there were in the 1950s. Proper documentation of them would overbalance what is really just a research note but would include at the least the following:

- Affect control theory
- Affect exchange theory
- Power and bargaining theory
- Behavioral exchange theory
- Critical mass theory
- Elementary theory
- E-state structuralism
- The game theory of power
- At least two programs of theory and research on social dilemmas
- Identity control theory
- Justice theory
- Social identity theory
- Social influence network theory
- The theory of collective action
- The theory of legitimate authority

The proper documentation of at least these examples can be found in one or more of [Berger and Zelditch \(1993, 2002\)](#), [Berger, Willer, and Zelditch \(2005\)](#), [Burke \(2006\)](#), and Sell and Reese (Chapter 10, this volume).

A theoretical research program can grow to be quite a complex structure, but it is complex in a different way than an effect program. In the first instance, of

---

4. For a more complete account, see [Wagner and Berger \(2002\)](#).

course, however complex, it is still about one process, not many processes. The different theories are theories of the same process under different conditions and not, as most effect programs, different processes under the same conditions. However, it is also a more coherent structure because, like expectation states theory, all of them have a common core, like the concept of an expectation state, that interrelates their parts.<sup>5</sup> This makes a considerable difference to how they grow, how much they grow, and the impact of how they grow.

## VI ASSESSMENT

Not only were there more of them but also theoretically oriented experiments had more impact than effect experiments, and theoretical research programs have shown more growth.<sup>6</sup> This was due partly to their nature and partly to their functions. In the first instance, because theories are analytic simplifications of some concrete phenomenon, and hence they deal with only one process, theoretically oriented experiments are also analytic simplifications of some concrete phenomenon, also dealing with only one process. This in itself makes an important difference to the impact of experiments and coherence of the programs that emerge out of them. But in addition to the nature of the experiments in them, programs of them are interrelated networks, not only of theories and models but also of theorists and researchers. The network of theories and models is interrelated by a core of strategies, concepts, propositions, and methods common to all of them. The networks of theorists and researchers are interrelated by not only interpersonal ties but also a shared background of norms and standards, theory and research. The effect is like that of a paradigm in Kuhn (1962): the existing state of the art—the program’s theories, the theoretical research in support of them, and the applied research using them—defines for its network of theorists and researchers what problems have or have not been solved; which of the unsolved problems are or are not theoretically significant—that is, which will matter more to growth of its theories; of the theoretically significant unsolved problems, which are or are not solvable, both in the sense that it promises a solution actually exists and makes available the conceptual and methodological tools it requires; and finally, its standards define whether or not a proposed solution is acceptable (Berger et al., 2005).

In consequence, theoretical research programs focus on, and invest their resources in, theory and research on unsolved problems that are theoretically significant and solvable. This is true no matter what the program’s methods of observation and inference, but if its method is experimental, the standardization of its basic experimental setting, characteristic of such programs, facilitates the

---

5. For a more complete account, see Wagner and Berger (2002).

6. Not all experiments, theoretical or not, have the impact that their nature and functions make possible. Nor do all programs, theoretical or not, achieve the growth their nature and functions make possible. Some fail. But whether or not they succeed, they all have greater potential impact than an effect experiment and greater capacity for theoretical growth than an effect program.

cumulative impact of any experiment done by the program.<sup>7</sup> Because its output is theoretically significant, the theory it is oriented to is a unit process, and the theories of the program are interrelated by a core of elements common to all of them: it is coherent, interrelated with both what has gone before and what comes after it. Because it is not only a network of theories but also a network of theorists and researchers—who share the same theory and research, the same sense of what is significant, of what is solvable—it focuses the theoretical research of an entire community on addressing the problem. Because a community, not just one isolated theorist, is involved, more—and more various—solutions of them are proposed. Because a solution is promised, more resources are invested in searching for it, even if finding it is likely to be deferred, and the entire network persists in a sustained search for a solution. In fact, because there remain other unsolved problems even if a solution is found, the entire career of a theoretical research program is one sustained, continually evolving, body of theory and research. Finally, not least important, because the network shares common norms, values, and standards of assessment, disagreements over the interpretation of solutions focus on conflicts reconcilable by appeal to reason and evidence. Hence, experiments in theoretical research programs have a greater cumulative impact than other experiments, and theoretical research programs have a greater capacity for growth—and sustain it for longer periods of time—than effect programs, which tend either to peter out, as the group decisions program did, or become incoherent, as the Asch program did, or morph into theoretical research programs, as the Bales program did.

They also have a greater capacity for the growth of applied research than effect programs. One feature of effect experiments and programs that I have not dwelled on up to this point is that because their effects are underanalyzed, their results are not generalizable. The problem of generalizing from one concrete instance to another depends on a theory: it requires definition of what constitutes an instance of the phenomenon and of its scope—that is, the conditions under which a result is or is not applicable. Effect experiments and programs, because they do not define the instantiation and scope conditions of their effects, leave the question of their generalizability unanswered. Theoretically oriented experiments, because their instantiation and scope conditions are defined by the theories they test, make such generalization possible. Hence, they successfully bridge the gap between experiments that test the theory and use of the theory to predict or explain the phenomena to which the theory is applicable in natural situations. The applied research that is integrated into the program is every bit as important as its theoretical research, and it gives to theoretical research programs an explanatory power of which no effect program is capable.

---

7. Effect programs, in which replications are frequent, also standardize experimental settings. The advantage of standardization, for either type of program, is that it makes it possible to compare and contrast the effects of different conditions between experiments with the same confidence as within experiments. What makes the standardized experiment of a theoretical research program more effective is that it standardizes an analytic simplification of one process, focusing only on oranges instead of trying to compare the oranges with apples.

## VII CHALLENGE: THE EXTERNAL VALIDITY OF EXPERIMENTS THAT TEST THEORIES

### A The Problem

Because they randomize the allocation of subjects to treatments, experiments solve the problem of internal validity, the validity of causal inference, better than any other method. However, there are many features of an experiment that it does not randomize: its simplifications, its artifices, and its population, period, and setting are all fixed initial conditions. If any of the fixed initial conditions specific to its methods interact with the hypothesis it tests, it may be internally valid but it may not be valid to generalize it. Common wisdom has it that experiments trade external validity, their generalizability, for internal validity. Because there is no way to randomize a constant, randomization does nothing to solve the problem of external validity. However, no trade-off is acceptable if an experiment is theoretically oriented. If its purpose is to test, refine, or extend a theory, an experiment that is not generalizable is fatally flawed.

### B The External Validity of Effect-Oriented Experiments

Campbell and Stanley (1963), the earliest systematic investigators of external validity, asked, more or less atheoretically, to what other population, setting, treatments, or measures could the effect of an experiment be generalized. Absent theory, the answer appeared to depend on how concretely similar the experiment was to the natural setting to which it was generalized. But even effect-oriented experiments seldom look like any concrete natural setting. Theory-driven experiments look even less like them. Their simplifications, artifices, manipulations and controls, and frequent enhancements of a process in ways seldom seen in nature (e.g., repeated trials) are sources of numerous concrete dissimilarities between the experiment and any natural setting to which it is generalized.

However, neither concrete similarities nor dissimilarities between the two make any difference to the generalizability of the experiment. If everything were relevant to everything else, there would be no generalization. But not everything in its natural setting is relevant to the theory being tested by an experiment. Concrete dissimilarities matter only if they are theoretically relevant. One does not say that an experiment is externally invalid because it controls air resistance or friction. But the concrete similarity of variables that are theoretically relevant does not matter any more than the concrete dissimilarity of those that are not.

It does matter that the process in an experiment and the situation to which it is generalized are both instances of the theory tested by the experiment. But theories are (by definition) capable of multiple concrete interpretations. What matters is theoretical, not concrete, similarity. The fictions common to many experiments, for example, do not matter to the generalizability of the experiment if they have the same meaning to the Ss in the experiment that their counterparts would have in the natural setting to which the experiment is generalized. If the

experiment is made up of all and only the variables relevant to the theory it tests and they all have the same meaning in the experiment that they would have in its natural setting, then the experiment has “experimental,” even if not “mundane,” realism. It is experimental, not mundane, realism that generalization from experiments requires ([Berkowitz & Donnerstein, 1982](#)).

## C The External Validity of Experiments That Test Theories

But what generalizes is the theory—not the experiment, not its effect, not even its hypothesis. Quite apart from the fact that, taken as a whole, a concrete experiment is as unique as a snowflake, hence, absent theory, ungeneralizable—or that, taken by itself, an effect is only conditional, hence ungeneralizable absent conditions—even an abstract, conditional hypothesis, taken by itself, can almost never be directly extrapolated because almost never does a single experiment test the whole of a theory. More often, any particular hypothesis controls some factors in order to test the effects of others. The fundamental test of an experiment’s generalizability is application of the theory it tests to some natural setting to which the theory applies. But natural settings seldom control the variables controlled by the experiment. What predicts and explains the behavior of a process in any natural setting is the theory as a whole, not any single test of it. If a test supports a theory, the hypothesis it tests does, of course, hold in any natural setting to which the theory applies. But its effect is complicated by the effects of other factors in the theory. A particular test is related to application of the theory it tests only indirectly, through the theory it supports, not by directly extrapolating it.

Absent theory, the question [Campbell and Stanley \(1963\)](#) asked is virtually unanswerable. But if an experiment tests a theory, if both experiment and natural setting are instances of it, a necessary condition of the external validity of the experiment is that no method specific to it interacts with its theory but is constant in the experiment. This is as true of what it does not do as what it does do. Hence, another way to put it is that an experiment is externally valid if it is an instance of the theory it tests and neither adds anything to nor subtracts anything from the theory it tests.

## D Advances in Detecting, Diagnosing, Testing, and Remedyng the External Invalidity of Experiments That Test Theories

If a theory tested by an experiment does not generalize and one suspected cause is the external invalidity of the experiment,<sup>8</sup> much effort goes into detecting its sources. However, diagnosis of a source is not, by itself, sufficient. One must still test the hypothesis that the suspected threat actually accounted for the external invalidity of the experiment. Furthermore, even if the threat did interact with

---

8. Other suspected causes are flaws in the theory, its measurement, and its application ([Zelditch, 2007](#)).

the treatments of the experiment, it does not follow that absent the threat the theory would generalize. All a test of a diagnosis proves is that the experiment did not actually test the theory. One must still retest the theory, either eliminating the threat, controlling for it, or estimating the magnitude of its effect.

The threats themselves are basically of three kinds: errors of instantiation—that is, failure of the experiment to satisfy the instantiation and scope conditions of the theory it tests; errors of commission—that is, interactions not in the theory but introduced into the experiment by its methods; and errors of omission—that is, interactions in the theory but omitted from the experiment by its methods.<sup>9</sup> Both instantiation and scope conditions are problematic: status characteristic experiments cannot take for granted that gender will always be a status characteristic (Foschi & LaPointe, 2002) or normative influence experiments that an aggregate will always be a group (Deutsch & Gerard, 1955). But even if one satisfies the instantiation and scope conditions of a theory, the social psychology of the experiment threatens its external validity: simply awareness that one is a subject in an experiment, measurement effects, demand characteristics, experimenter effects, subject voluntarism and expectations, especially evaluation apprehension and social desirability effects, all threaten its external validity (Rosenthal & Rosnow, 1969). Furthermore, what one does not do is as much a threat as what one does do. Ineffective manipulation of independent variables (e.g., artifice that is not real to Ss) or moderators (e.g., incentives), or restriction of their range (e.g., time horizons), especially if any of them are nonlinear, may equally undermine external validity (Zelditch, 2007).

But developments in both the design and the analysis of experiments have improved methods of detecting, diagnosing, testing, and remedying flaws in their external validity. Although some experimentalists have been hostile to it, causal modeling of experiments (Costner, 1971), use of it to detect measurement and specification errors, and the confirmatory meta-analysis made possible by it (Hedges & Olkin, 1985)<sup>10</sup> have advanced methods of detecting and diagnosing flaws in the external validity of experiments that test theories.

Experiments themselves have played a crucial role in detecting flaws such as the Hawthorne effect (Roethlisberger & Dickson, 1939) and demand

9. I omit errors of application. Applied research is as important to the external validity of experiments as theoretical research. Experiments that test a theory are a necessary but not sufficient condition of its application. Prediction or explanation of its behavior in any application depends not only on theoretical research supporting the theory but also on tests that show that the situation to which it is applied satisfies its instantiation and scope conditions, field research that describes the initial conditions of the principles of the theory in the particular situation and any information available about any other processes in the situation that interact with them—all empirical questions independent of the theoretical research in support of the theory. Applied research is as vulnerable to the extraneities of its methods as theoretical research, but they are independent of its theoretical research (Zelditch, 2007).

10. Confirmatory meta-analysis estimates heterogeneity due to method of different methods of testing the same hypothesis.

characteristics (Orne, 1962). But developments in research design have also advanced methods of testing and remedying them. The control group has been extended to the social psychology of the experiment. Separating an experiment's social psychology from its treatment effects, reactivity control groups have evolved to control for measurement effects, demand characteristics, experimenter expectations, evaluation apprehension, the voluntarism of Ss, subject expectancies, and subject awareness of the experiment (e.g., see Rosenthal, 1976). Simulation of experiments has used an observer not exposed to an experiment's treatments to the same effect (e.g., Alexander & Knight, 1971). In turn, both have led to the altered replication of flawed tests of a theory that uses them to remedy its flaws. It may appear illogical to argue that a flawed method is, after all, the best method of testing and remedying the flaws in the external validity of experiments that test theories. But it has been experiments that have been crucial in demonstrating the flaws in experiments and experiments that have proved most useful in testing and remedying them.

## REFERENCES

- Alexander, C. N., & Knight, G. W. (1971). Situated identities and social psychological experimentation. *Sociometry*, 34, 65–82.
- Allen, V. L. (1965). Situational factors in conformity. *Advances in Experimental Social Psychology*, 2, 133–175.
- Allen, V. L. (1977). Social support for nonconformity. *Advances in Experimental Social Psychology*, 8, 1–43.
- Asch, S. E. (1951). Effects of group pressure upon the modification and distortion of judgments. In H. Guetzkow (Ed.), *Groups, leadership, and men* (pp. 177–190). Pittsburgh, PA: Carnegie Press.
- Back, K. (1951). Influence through social communication. *Journal of Abnormal and Social Psychology*, 46, 9–23.
- Bales, R. F. (1953). The equilibrium problem in small groups. In T. Parsons, R. F. Bales, & E. H. Shils (Eds.), *Working papers in the theory of action* (pp. 111–161). Glencoe, IL: Free Press.
- Bales, R. F., & Slater, P. (1955). Role differentiation in small decision making groups. In T. Parsons, & R. F. Bales (Eds.), *Family, socialization and interaction process* (pp. 259–306). Glencoe, IL: Free Press.
- Bales, R. F., Strodtbeck, F. L., Mills, T. M., & Rosebourough, M. E. (1951). Channels of communication in small groups. *American Sociological Review*, 16, 461–468.
- Bavelas, A. (1950). Communication patterns in task-oriented groups. *Journal of the Acoustical Society of America*, 22, 725–730.
- Bennett, E. B. (1955). Discussion, decision, commitment, and consensus in group decision. *Human Relations*, 8, 251–274.
- Berger, J. (1958). Relations between performance, reward, and action opportunities in small groups. PhD dissertation, Cambridge, MA: Harvard University.
- Berger, J., Cohen, B. P., & Zelditch, M. (1966). Status characteristics and expectation states. In J. Berger, M. Zelditch, & B. Anderson (Eds.), *Sociological theories in progress: Vol. 1*. (pp. 29–46). Boston: Houghton Mifflin.
- Berger, J., & Conner, T. L. (1969). Performance expectations and behavior in small groups. *Acta Sociologica*, 12, 186–198.

- Berger, J., Fisek, M. H., Norman, R. Z., & Zelditch, M. (1977). Status characteristics and social interaction: An expectation-states approach. New York: Elsevier.
- Berger, J., Willer, D., & Zelditch, M. (2005). Theory programs and theoretical problems. *Sociological Theory*, 23, 127–155.
- Berger, J. & Zelditch, M. (Eds.). (1993). *Theoretical research programs: Studies in the growth of theory*. Stanford, CA: Stanford University Press.
- Berger, J., & Zelditch, M. (1997). Theoretical research programs: A reformulation. In J. Szmata, J. Skvoretz, & J. Berger (Eds.). *Status, network, and structure: Theory development in group processes* (pp. 29–46). Stanford, CA: Stanford University Press.
- Berger, J. & Zelditch, M. (Eds.), (2002). *New directions in contemporary sociological theory*. New York: Rowman & Littlefield.
- Berkowitz, L., & Donnerstein, E. (1982). External validity is more than skin deep: Some answers to criticisms of laboratory experiments. *American Psychologist*, 37, 245–257.
- Burke, P. (1967). The development of task and social-emotional role differentiation. *Sociometry*, 30, 379–392.
- Burke, P. (Ed.). (2006). *Contemporary social psychological theories*. Stanford, CA: Stanford University Press.
- Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research. Chicago: Rand McNally.
- Costner, H. (1971). Utilizing causal models to discover flaws in experiments. *Sociometry*, 34, 398–410.
- Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. *Journal of Abnormal and Social Psychology*, 51, 629–636.
- Emerson, R. M. (1964). Power-dependence relations: Two experiments. *Sociometry*, 27, 282–298.
- Faucheu, C., & Mackenzie, K. D. (1966). Task dependency of organizational centrality: Its behavioral consequences. *Journal of Experimental Social Psychology*, 2, 361–375.
- Festinger, L. (1950). Informal social communication. *Psychological Review*, 57, 271–282.
- Foschi, M., & Lapointe, V. (2002). On conditional hypotheses and gender as a status characteristic. *Social Psychology Quarterly*, 65, 146–162.
- Glanzer, M., & Glaser, R. (1961). Techniques for the study of group structure and behavior: II. Empirical studies of the effects of structure in small groups. *Psychological Bulletin*, 58, 1–27.
- Hedges, L. V., & Olkin, I. (1985). Statistical methods for meta-analysis. New York: Academic Press.
- Kuhn, T. S. (1962). The structure of scientific revolutions. Chicago: University of Chicago Press.
- Lewin, K. (1947). Group decision and social change. In T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology* (pp. 330–344). New York: Henry Holt.
- Lewis, G. H. (1972). Role differentiation. *American Sociological Review*, 37, 424–434.
- Mulder, M. (1960). Communication structure, decision structure, and group performance. *Sociometry*, 23, 1–14.
- Orne, M. T. (1962). On the social psychology of the psychology experiment. *American Psychologist*, 17, 776–783.
- Roethlisberger, F. J., & Dickson, W. J. (1939). Management and the worker. Cambridge, MA: Harvard University Press.
- Rosenthal, R. (1976). Experimenter effects in behavioral research (enlarged ed.). Cambridge, MA: Harvard University Press.
- Rosenthal, R., & Rosnow, R. L. (Eds.). (1969). *Artifact in behavioral research*. New York: Academic Press.
- Schachter, S. (1951). Deviation, rejection, and communication. *Journal of Abnormal and Social Psychology*, 46, 190–207.

- Sherif, M. (1935). A study of some social factors in perception. *Archives of Psychology*, 27(No. 187).
- Wagner, D. G., & Berger, J. (2002). Expectation states theory: An evolving research program. In J. Berger, & M. Zelditch (Eds.), *New directions in contemporary sociological theory* (pp. 41–76). New York: Rowman & Littlefield.
- Zelditch, M. (2007). The external validity of experiments that test theories. In M. Webster, & J. Sell (Eds.), *Laboratory experiments in the social sciences* (pp. 87–112). Burlington, MA: Elsevier.

## Chapter 9

# Experiments on Exchange Relations and Exchange Networks in Sociology

Linda D. Molm

*University of Arizona, Tucson, Arizona*

## I INTRODUCTION

This chapter reviews experimental research programs on social exchange and exchange networks as they have developed in sociology during the past 35 years. Many forms of interaction outside the economic sphere can be conceptualized as an exchange of benefits. Both social and economic exchanges are based on a fundamental feature of social life: much of what we need and value (e.g., goods, services, and companionship) can only be obtained from others. People depend on one another for these valued resources, and they provide them to each other through the process of exchange. Social exchange theories focus on this aspect of social life—the benefits that people obtain from, and contribute to, social interaction, and the patterns of interdependence that govern those exchanges. The social structures within which exchange takes place (exchange relations and networks), the different processes through which exchange occurs (e.g., bargaining, reciprocal gift-giving, and generalized exchange), and the behavioral and affective outcomes of exchange (including power inequalities, coalition formation, commitment, and trust) are all addressed by contemporary exchange theories and experimental programs testing these theories.

In this chapter, I review the theoretical background and historical development of the field that led to the strong tradition of experimental research in social exchange. I describe features of the major standardized settings and designs for the study of exchange; discuss the role of technological advances in their development; and describe several experiments that illustrate the use of the standardized settings to test, modify, and extend exchange theories. I conclude with an assessment of the current state and future prospects for experimental research in this area.

## II BACKGROUND AND DEVELOPMENT

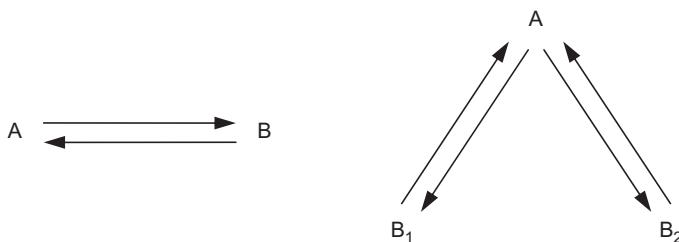
### A Basic Concepts and Assumptions

Social exchange occurs between two or more actors who are dependent on one another for valued outcomes. Social exchange theories assume that actors are motivated to obtain more of the outcomes that they value and others control, that actors provide each other with these valued benefits through some form of social exchange, and that exchanges between the same actors are recurring over time (rather than “one-shot” transactions). These scope assumptions are shared by most theories of exchange and must be met in the experimental settings in which they are tested.

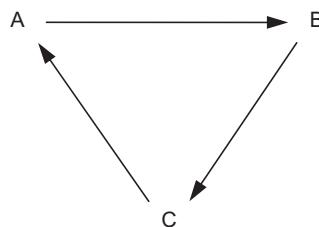
The simplest form of social exchange involves just two actors, *A* and *B*, each of whom possesses at least one resource that the other values. The actors can be either individuals or corporate groups (e.g., organizations), and the resources can include not only tangible goods and services but also capacities to provide socially valued outcomes such as approval or status. In exchange experiments, the actors are always individual persons, but sometimes they are given roles as representatives of organizations. Exchange theories make no assumptions about *what* actors value and assume that interaction is unaffected by actors’ values or the resources exchanged; this makes them broadly applicable to social relations regardless of content and means that experimental tests of exchange theories can use any resource of known value. Some exchange theories assume “rational” actors who cognitively weigh the potential benefits and costs of alternative partners and actions and make choices that maximize outcomes; others adopt a learning model that assumes actors respond to consequences of past choices, without conscious weighing of alternatives and without necessarily maximizing outcomes.

As [Figure 9.1](#) illustrates, social exchange can take several distinct forms: direct exchange; generalized exchange; and productive exchange. In relations of *direct exchange* between two actors, each actor’s outcomes depend directly on another actor’s behaviors; that is, *A* provides value to *B*, and *B* to *A*, as in the example of two co-workers helping each other with various projects. As [Figure 9.1a](#) shows, such direct exchange relations can occur either in isolated dyads or within larger networks. In relations of *generalized exchange* among three or more actors, each actor gives benefits to another and eventually receives benefits from another, but not from the same actor. Consequently, the reciprocal dependence is indirect; a benefit received by *B* from *A* is not reciprocated directly by *B*’s giving to *A* but, rather, indirectly by *B*’s giving to another actor in the network. Some forms of indirect exchange (e.g., the classic Kula ring) take a specific circular form, as shown in [Figure 9.1b](#). Other examples, such as donating blood and reviewing journal manuscripts, do not. Finally, in *productive exchange* ([Figure 9.1c](#)), both actors in the relation must contribute in order for either to obtain benefits (e.g., co-authoring a book).

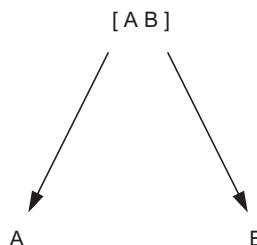
Although generalized exchange was a particular interest of early anthropological exchange theorists, the study of direct exchange relations has dominated



(a) Direct Exchange



(b) Generalized (Indirect) Exchange



(c) Productive Exchange

**FIGURE 9.1** Direct, generalized, and productive exchange structures.

research and theorizing in sociology until quite recently. Direct exchanges can be *negotiated* or *reciprocal* in form; both have been the subject of long-term research programs. In negotiated exchange, actors jointly negotiate the terms of an agreement (usually binding) through a series of offers and counteroffers. Each agreement comprises a discrete transaction that provides benefits for both actors. In reciprocal exchange, actors perform individual acts that benefit another, such as giving assistance or advice, without negotiation and without

knowing whether, when, or to what extent the other will reciprocate. Exchange relations develop when beneficial acts prompt reciprocal benefit.

## B Historical Development

### 1 The Early Theories

The development of the exchange perspective in sociology began in approximately 1960, with the publication of theories by George Homans (1961), Peter Blau (1964), and John Thibaut and Harold Kelley (1959). The experimental research tradition, in contrast, did not begin until the late 1970s. The early works of Homans and Blau demonstrated the ubiquity of exchange processes in social life and stimulated interest in (and controversy over) the perspective, but their narrative style was not conducive to experimental test. Thibaut and Kelley (both psychologists) provided a more analytical framework for the study of social exchange and introduced a tool for describing patterns of interdependence in relations—the outcome matrix, which was used in experimental research throughout the 1960s and 1970s. Although its usefulness was largely restricted to dyads, this work provided the impetus for later, more sophisticated studies of social exchange.

### 2 Emerson's Power-Dependence Theory

The contemporary tradition of social exchange and the rise of programmatic experimental research on exchange began with the publication of Richard Emerson's exchange formulation in 1972 (1972a, 1972b). Emerson's theory, which built upon his earlier (1962) work on power-dependence relations (and, consequently, came to be called power-dependence theory), influenced the development of theory and research on exchange in several important ways. First, his rigorously derived system of propositions moved the tradition toward a more formal, analytical approach to theory that was more amenable to experimental test. Second, Emerson made the structure of relations, rather than the actors themselves, the central focus of his theory, and exchange networks replaced dyadic relations as the primary structural unit. Third, the theory established power and its use as the major topics of exchange theory—topics that would dominate research on social exchange until the late 1990s. Finally, the research program that Emerson, Karen Cook, and their students began in the late 1970s shifted the focus from the reciprocal exchanges of the classical exchange theorists to negotiated exchanges in which actors bargained over the terms of agreements.

One of Emerson's most influential contributions was the concept of *exchange networks*, defined as sets of direct exchange relations that are *connected* to one another. Connected relations are linked by a focal actor (e.g., A–B–C), and exchange in one relation (e.g., the A–B relation) affects the frequency or value of exchange in the other (the B–C relation). Connections are *positive* to the extent that exchange in one relation increases exchange in the other (e.g., B's exchange with A might give B a resource that B can then use in exchange with C), and they

are *negative* to the extent that exchange in one relation decreases exchange in the other.

Negatively connected relations form the structural foundation for Emerson's theory of power-dependence and for most theories of power in exchange networks that were subsequently developed.<sup>1</sup> Negative connections provide actors with *alternative* exchange partners who compete with one another for the opportunity to exchange with the focal actor. For example, in the negatively connected network  $B_1$ -A- $B_2$ , A might be an employer and the  $B$ s applicants for the same job, or the  $B$ s might be potential tennis partners for A. Alternative partners decrease an actor's dependence on any one exchange partner, giving that actor an advantage in exchanges with more dependent partners. Alternative partners can vary in both their *value* and their *availability* as an exchange partner. The former is a function of the value to an actor of the resources controlled by the partner, and the latter is a function of the size and shape of the exchange network. The more valuable and the more available A's alternative to B is, the less dependent A is on B.

Emerson proposed that power in the A-B relation can be described by two dimensions: the power balance or imbalance in the relation (i.e., actors' *relative* power over each other, determined by the difference in their dependencies) and the cohesion in the relation (actors' *absolute* power over each other, determined by the sum of their dependencies). Later theories focused almost exclusively on relative power. Over time, the structure of power produces predictable effects on the frequency and distribution of exchange benefits. The more dependent actors are on each other, the more frequently they exchange with each other; the more imbalanced (unequal) their power dependencies are, the more unequal their exchange is, with the less dependent, more powerful actor receiving more benefits at lower cost.

### 3 The Emergence of New Theories of Power in Networks

With the concept of exchange networks, theory and research on social exchange shifted from the study of dyadic relations to the larger opportunity structures within which those relations are embedded. How network structures affect the availability of alternatives and how the value and availability of alternatives affect the distribution of power became the central focus of experimental research on exchange networks during the 1980s and 1990s. Richard Emerson and Karen Cook conducted the first experimental tests of Emerson's power-dependence theory in the late 1970s and early 1980s, showing that power use—the asymmetry in exchange benefits between A and B—can be predicted from the structural power relations in exchange networks. As their research program developed, however, it became apparent that Emerson's formulation was inadequate for analyzing

---

1. In the network exchange tradition developed by Barry Markovsky and David Willer, negative connections are called exclusive connections (Szmata & Willer, 1995).

power relations in more complex networks (Cook, Emerson, Gillmore, & Yamagishi, 1983). As a result, new theories of network exchange and power emerged (including a reformulation of power-dependence theory), with varying algorithms capable of predicting the distribution of power in networks as a whole.

The major competitor to power-dependence theory to emerge, and the one that has produced the largest body of experimental research, is the network exchange theory (NET) of Barry Markovsky, David Willer, and associates (Markovsky, Willer, & Patton, 1988). This theory draws on Willer's elementary theory; later, it was modified to incorporate a resistance model. Other approaches include Friedkin's (1992) expected value theory and Bienenstock and Bonacich's (1992) application of core theory. All of these theories use formal mathematical models to predict the distribution of power in exchange networks, and all focus primarily on negotiated exchanges in negatively connected networks that vary in size, shape, and complexity.

Network exchange theory uses a path-counting algorithm called the graph-theoretic power index (GPI) to predict the relative power of each position in a network; positions with higher GPI scores are predicted to receive a larger share of the profit from agreements. Like power-dependence theory, NET assumes structural power is derived from the availability of alternative partners. Alternatives are more available to an actor if they lie on odd- rather than even-length paths (i.e., alternatives on odd-length paths have no alternatives themselves, or their alternatives are less available because they have other alternatives). Differences in availability affect the likelihood that some actors will be excluded from exchange on each opportunity; exclusion increases power use by driving up the offers that excluded actors make on subsequent negotiation opportunities.

#### *4 Other Questions, New Directions*

At the same time that research on power in exchange networks was developing, other research programs, particularly one headed by Edward Lawler and a second directed by Linda Molm, were studying somewhat different exchange questions: how dynamic processes of power use (and not only structure) affect exchange patterns and how the capacity and use of both reward-based and punitive power affect exchange. Both Lawler and Molm argued that power use can be strategic as well as structurally induced, and both studied power that was based not only on control over rewards but also on control over punishments. Lawler's (1992) work continued the focus of exchange researchers on bargaining relations (negotiated exchanges), whereas Molm's (1981, 1997) work introduced the study of reciprocal exchanges in which actors provide benefits to each other without bargaining or negotiation. Such exchanges, which were the focus of the classical exchange theorists, are arguably more common in social life than the negotiated exchanges that most contemporary researchers have studied.

In their work on power, both Lawler and Molm continued to emphasize structure as the dominant force on exchange. However, they also brought in

consideration of other factors: cognitions, affect, risk, and fairness. This work set the stage for the development of new directions in exchange theories and research.

Beginning in the 1990s and continuing to the present, exchange theorists began to pay renewed attention to some of the long-neglected concerns of the classical theorists: the risk and uncertainty inherent in exchange (particularly generalized exchange and reciprocal exchange); the emergence of affective ties between exchange partners and their ability to transform the structure and nature of exchange; and the effects of different forms of exchange—negotiated, reciprocal, generalized, and productive—on the development of trust, commitment, and affective attachments. Lawler and colleagues developed theories of relational cohesion, affect, and commitment in social exchange that emphasized the causal role of emotions in social exchange processes; Molm and colleagues began systematic comparisons of different forms of exchange that examined how the structure of reciprocity in exchange affects the development of trust and solidarity; and Peter Kollock, Toshio Yamagishi, and Karen Cook studied the relations among uncertainty, trust, and commitment.<sup>2</sup>

This work shifted theories and research in the social exchange tradition in two important ways: first, toward consideration of forms of exchange other than direct negotiated exchange; and, second, to the study of integrative outcomes (trust, commitment, and affective ties) rather than the differentiating outcomes (power and inequality) that had dominated exchange research. These changes also led to the development of new experimental settings, particularly for the study of different forms of exchange.

### III STANDARD SETTINGS AND DESIGNS

Because most theories of exchange are concerned with basic causal processes linking various aspects of exchange, most of the research testing these theories uses experimental methods and standardized laboratory settings. Laboratory experiments have several advantages for studying exchange and for testing theories formulated at the abstract level of most exchange theories.

As we have seen, most contemporary theories of exchange are concerned, in some way, with the effects of exchange *structures*. In natural settings, it can be very difficult to separate the effects of characteristics of structures from the effects of characteristics of actors who occupy particular positions in the structures. In laboratory experiments, random assignment of subjects to positions in exchange relations or networks accomplishes this task. It can also be difficult in natural settings to measure or compare exchange relations that are based on

---

2. For examples of work on these topics, see Kollock (1994), Lawler (2001), Lawler and Yoon (1996), Lawler, Thye, and Yoon (2008, 2009), Molm, Collett, and Schaefer (2006, 2007), Molm, Takahashi, and Peterson (2000, 2003), Yamagishi and Cook (1993), and Yamagishi, Cook, and Watabe (1998).

different resources (e.g., approval, expert advice, or status); many exchange resources cannot easily be quantified, and their value varies across different actors and changes over time. In laboratory experiments, a single, widely valued, quantifiable resource is typically used: money.

In developing a laboratory setting for the study of social exchange, several considerations are important. First, the setting must meet the general scope conditions of the theory. For most exchange theories, this means that the setting must: (1) give actors control over resources that provide outcomes of value to other actors; (2) structure exchange relations and networks that create mutual dependencies among actors for those outcomes; and (3) provide repeated opportunities for exchange among the same actors (or among interchangeable occupants of the same positions). Second, the setting must make it possible to manipulate dimensions of the structure and process of exchange and measure their effects on exchange outcomes (e.g., power use, behavioral commitment, and affective ties) in ways that are consistent with the theory's conceptual definitions of these concepts. Third, variables that are unrelated to the theory and extraneous to the research must be either controlled or randomly distributed across the experimental conditions.

Several standardized laboratory settings have been created that meet these requirements in various ways. The first standardized setting for the study of social exchange networks was developed by [Karen Cook and Richard Emerson \(1978\)](#),<sup>3</sup> who created a setting for the study of negotiated exchange. Basic parameters of their setting were later adopted, with some notable differences, by researchers testing other theories of network exchange and power. [Samuel Bacharach and Edward Lawler \(1981\)](#) also developed a standardized setting for studying power in negotiated exchange, but in dyads rather than networks. [Linda Molm \(1981\)](#) created a setting for the study of reciprocal exchange and later (1997) extended it to include both reward-based and coercive exchanges. More recently, several researchers have created settings for studying generalized and productive exchange ([Lawler et al., 2008](#); [Molm et al., 2007](#); [Yamagishi & Cook, 1993](#)), although a standardized setting has yet to emerge.

All of these settings share certain features. Subjects are randomly assigned to positions in a particular network or dyadic structure (sometimes varied as one of the factors in the experiment), and their exchange behavior within that structure is studied. The network creates an opportunity structure for exchange and determines the actors in the network with whom each subject can exchange. Variations in the size, shape, types of connections, and potential value of exchange relations determine actors' absolute and relative power. To ensure that behavior is affected solely by manipulated (typically structural) characteristics of the exchange relations or networks and not by actors' personal characteristics,

---

3. An earlier setting was developed by [John Stolte and Richard Emerson \(1977\)](#); however, it was never established as a standardized setting used in multiple experiments.

subjects typically interact with one another through computers rather than face-to-face. They engage in repeated exchanges with other actors in the network, varying in number from 20 to several hundred in different experiments and different research programs.

The benefits that subjects receive from these exchanges are operationalized as points, equal to money; subjects earn money through exchanges and at the conclusion of the experiment are paid what they earn.<sup>4</sup> To meet the assumption of self-interested actors, subjects are recruited on the basis of their interest in earning money. Because no other individual characteristics are theoretically relevant, most researchers take advantage of the convenience of undergraduate students as the potential pool of subjects.

The following sections describe the specific features of the settings developed to study the different forms of exchange: negotiated, reciprocal, generalized, and productive.

## A Negotiated Exchange Setting

In the original [Cook and Emerson \(1978\)](#) setting, subjects negotiated the terms of exchange, through a series of offers and counteroffers, to reach binding agreements on each of a series of negotiation opportunities. The agreements reached determined how many points—equal to money—each subject received. During each transaction period, subjects could send offers and counteroffers to any of the partners to whom they were connected in their network. If an agreement was reached during that time, both subjects received the benefits they had agreed upon; if no agreement was reached before the end of the transaction period, no benefits were received on that opportunity.<sup>5</sup> To eliminate effects of equity concerns on exchange agreements, subjects had no knowledge of the benefits their partners received from their agreements. In reality, subjects divided a fixed amount of profit between them, but they were unaware of the division, the total profit, or their partner's gain.

The Cook and Emerson setting (like most others that followed) was specifically designed for the study of power in negatively connected networks.<sup>6</sup>

---

4. Money is used as the valued benefit because of its advantages for experimental control: money is widely valued; it can be quantified to produce a ratio level of measurement; and it is resistant to the effects of satiation or diminishing marginal utility (which would alter the value of the resource in unknown ways). The exchange resource in the experiments is not money per se—that is, money is not transferred from one actor to another, as in economic exchanges—but, rather, the capacity to produce valued outcomes, operationalized as money, for another.

5. In the Cook and Emerson setting, each transaction period was of a specified length of time; in other negotiated exchange settings, a specified number of rounds of offers and counteroffers is allowed. For example, in [Lawler and Yoon's \(1996\)](#) negotiated exchange setting (which Molm and colleagues adopted with minor modifications), subjects had five rounds to make an agreement.

6. For studies of positively connected networks, see [Yamagishi, Gillmore, and Cook \(1988\)](#) and [Schaefer and Kornienko \(2009\)](#). Related distinctions in the network exchange tradition between exclusive and inclusive connections have been studied by [Szmata and Willer \(1995\)](#).

To create the negative connections, a subject's exchange with one partner precluded exchange with another partner on that opportunity. Thus, each actor in the network could make only one agreement per opportunity. The setting was also designed to test Emerson's assertion that power leads to power use regardless of actors' knowledge or intentions; accordingly, subjects were not informed of the size or shape of the network beyond their immediate connections. This meant that actors were unaware of any power advantage (or disadvantage) that they or others in the network enjoyed.

As competitors to power-dependence theory emerged, including Markovsky and Willer's network exchange theory, Friedkin's expected value theory, and Bienenstock and Bonacich's core theory, these researchers adopted the basic parameters of the Cook and Emerson setting, particularly the process of having subjects bargain over the division of a fixed pool of profit points. Some important modifications were made, however. In the original experiments testing network exchange theory (e.g., [Markovsky et al., 1988](#)), subjects were given full information of the network structure, the points each actor received from agreements, and the earnings of all positions in the network. Rather than using restricted information to control effects of subjects' equity concerns, the network exchange setting rotated subjects through all power positions in a network. The rationale for this procedure (as a control for equity effects) is that because all subjects know they will be in both high- and low-power positions at some point, all should try to maximize their earnings in each power position rather than striving for equal agreements. From an experimental standpoint, however, the rotation is a within-subject variable with potential order effects; that is, the order in which subjects occupy the different positions may affect their behavior. Later experiments in this tradition returned to the restricted information procedures of Cook and Emerson, limiting information about the network and other subjects' earnings to control equity effects rather than rotating subjects through positions (e.g., [Thye, Lovaglia, & Markovsky, 1997](#)).

The network exchange tradition also treated the Cook and Emerson procedure for operationalizing negative connections—restricting the number of exchanges that a subject could make on each negotiation opportunity—as a variable condition of exchange. Varying this number can change the network connections, depending on the size of the network relative to the number of exchanges. As long as the number of exchanges that an actor can make on an opportunity is *less* than the number of potential partners with whom the actor can exchange, the relations that connect that actor to those partners will be negatively connected, and the actor's exchange with one partner will decrease the probability of exchange with another partner. But if the number of exchanges that an actor can make on an opportunity *equals* or *exceeds* the number of potential exchange partners for that actor, then the actor's exchange with one partner will have no effect on his or her exchange with other partners; the actor can exchange with *all* partners on each opportunity.

## B Reciprocal Exchange Setting

In Molm's (1981, 1997) reciprocal exchange setting, subjects exchange by individually performing acts—adding to or subtracting from a partner's points—that have rewarding or punishing consequences for the partners. In the standard setting, the points are of fixed value, which means that variations in the equality or inequality of exchange can occur only over time, based on subjects' relative rates of giving to one another.

On each of a series of exchange opportunities, subjects choose which partner to give points to (or, if both rewarding and punishing actions are possible, subjects choose both a partner and an action). Initiating exchange with one partner precludes initiating exchange with an alternative partner on that opportunity, thus creating negative connections between exchange relations. All subjects in the network make these choices simultaneously, without knowing in advance whether or when the target partner will reciprocate. They are then informed of their partners' behaviors: whether each of their potential exchange partners added to (or subtracted from) their earnings or did not act toward them.

Thus, on any given exchange opportunity, a subject might give to another without receiving, receive without giving, or reciprocally give and receive. Reflecting the learning model on which this exchange setting is based, subjects typically engage in exchanges for several hundred opportunities—far more than in negotiated exchange experiments. In the absence of explicit bargaining or knowledge of others' intentions, subjects can influence one another by making their behavioral choices contingent on their partners' previous choices.

Like the Cook and Emerson setting for negotiated exchange, the Molm setting for reciprocal exchange typically gives subjects only limited information about the network structure (i.e., subjects know only their immediate connections) and no information about the value of exchange benefits to their partners or their partners' cumulative earnings. The exception is experiments that are explicitly designed to study perceptions of fairness; then, subjects are given the full information necessary for making fairness judgments.

Other variations in the standard reciprocal exchange setting have included changing the amount given to the partner from a fixed value to a variable value, chosen by the subject from a range of points, and allowing subjects to keep any points not given to the partner (Lawler et al., 2008; Molm et al., 2006). In the latter variation, points given to another triple in value, whereas points kept for self remain the same. These variations introduce costs other than the opportunity costs that are part of the standard setting. They also illustrate, more generally, the use of systematic variation in experimental parameters within a standardized setting. Such variations are useful for identifying which aspects of a variable are responsible for producing particular effects and for ruling out alternative explanations of findings.

## C Generalized Exchange and Productive Exchange Settings

Generalized and productive exchange settings, both of which involve collective systems of exchange rather than the dyadic relations of direct exchange, are less developed. Fewer sociologists have studied these forms of exchange, and no true standardized settings (i.e., settings used across multiple experiments) have been developed. Yamagishi and Cook (1993) created a chain- or network-generalized exchange structure by giving participants a divisible resource (10¢) on each trial and allowing them to decide how much of this resource to give to the next person in the chain; points given to the other were doubled in value, whereas those kept for self remained the same. Molm and colleagues (2007) and Lawler and colleagues (2008) have developed generalized exchange settings as well. Molm operationalizes chain-generalized exchange as a variant of her reciprocal exchange setting, in which subjects chose whether or not to give a fixed number of points to their recipient in the chain (while receiving, or not receiving, points from their benefactor in the chain). Lawler's setting is similar to Molm's, but it includes a default benefit for subjects who choose not to give to their recipient.

Productive exchange has been studied by Lawler, Thye, and Yoon (2000, 2008). Their 2000 study operationalized it as a variant of Lawler and Yoon's (1996) negotiated exchange setting. Three subjects engaged in a joint venture, with each negotiating for a share of the profits produced by the joint venture; as in dyadic negotiations, subjects had to agree on a division of profit—in this case, a three-way division—before any could benefit. In their 2008 study, subjects instead decided individually whether or not to invest in a common account, which was then divided (after applying a multiplier) among all three subjects; in this operationalization, free riding was possible.<sup>7</sup>

## D Measurement of Dependent Variables

In the two direct exchange settings, behavioral measures include the frequency of exchange (measured by the frequency of agreements in negotiated exchange and the frequency of giving in reciprocal exchange), the inequality of exchange, behavioral commitment to particular partners, and (more rarely) the formation of coalitions of actors as a power-balancing strategy. In negotiated exchange, inequality (or power use) is measured by the agreed-upon division of the profit pool (i.e., the difference in points received by the two actors from their agreements). In reciprocal exchange, inequality is measured over time, by comparing the relative frequencies of rewarding behaviors that the two partners perform for one another. Exchange researchers also study the behavioral commitments that develop in particular dyadic relations within a network; commitment measures

---

7. Experiments on public goods (see Sell and Reese, Chapter 10, this volume) also involve a form of productive exchange in which free riding is possible; actors contribute to a collective good, which then benefits all members of the group.

typically compare the number of exchanges made with two or more alternative partners (Cook & Emerson, 1978; Kollock, 1994). Most exchange experiments do not allow actors to form coalitions; in the few that have, measures of which coalitions form and how often coalitions form have served as dependent variables (Cook & Gillmore, 1984).

Integrative outcomes such as trust, affective regard, perceptions of relational cohesion, and positive emotions are measured both through responses to semantic differential items asking subjects to evaluate their partners and relationships and through some behavioral tasks. Lawler and Yoon (1996), for example, measured commitment to partners by giving subjects the opportunity to give gifts to one another or to contribute to a joint venture. Responses to multiple semantic differential items are typically combined in scales to increase the reliability of the measure.

## E Designs and Design Issues

Most exchange experiments employ mixed between- and within-subject designs, typically varying such factors as network structure, relative power imbalance, or form of exchange “between subjects” (often in factorial designs, to test interactions) while treating actor positions within networks and trial blocks within the exchange period as “within-subject” variables.<sup>8</sup> Changes in behaviors over trial blocks are often studied to examine specific theoretical questions (e.g., whether power use increases or decreases over time) or to determine whether or when exchange patterns stabilize. Actor positions are compared to measure the inequality of exchange benefits (as a measure of power use) or to examine how actors who are relatively advantaged or disadvantaged on power differ in trust or feelings toward their partners.

The design employed also depends on the level of theory development. Theories that make ordinal predictions (i.e., predicting that mean values of a dependent variable will be higher in one condition than in another) typically use designs and methods of analysis that compare experimental conditions with each other. Theories that make specific point predictions instead compare expected values with observed values within experimental conditions.

Because both the actors within a relation or network and their transactions or exchange behaviors over time are interdependent, the unit of analysis for exchange experiments must be the relation or network and, for analyses of behaviors, some single measure of interaction for the exchange period must be used (e.g., mean behavior for the entire exchange period or for the last half of the exchange period). Recognition of these interdependencies is also theoretically consistent with the assumption of recurring exchanges between the same

---

8. Here, of course, the experimental terms “between subject” and “within subject” refer to between or within networks (or, in the case of trial blocks, the exchange period) because the experimental unit is the network rather than the individual subject.

actors and the development of *social exchange relationships*. In this respect, the network exchange tradition makes assumptions more typical of economic exchanges, often treating the transaction (rather than the exchange relation) as the unit of analysis and correcting for interdependencies through statistical means (Skvoretz & Willer, 1991).

## IV TECHNOLOGICAL DEVELOPMENTS

### A Computer-Mediated Exchange

Technological advances have shaped the development of exchange experiments during the past 35 years in important ways. First, computers have facilitated the high level of control that is a hallmark of laboratory experimentation. Computer-mediated interaction allows physically isolated participants to interact with one another without knowing (or being influenced by) characteristics of the other, such as gender, race, or age. It also allows researchers to structure the exchange process in particular ways, ensuring that all participants have the same kinds of information, interact for the same amount of time, make decisions and receive feedback in the same sequences, and so forth.

A 1991 study by John Skvoretz and David Willer compared a face-to-face exchange setting (used in the early research of Markovsky et al., 1988) with a computer-mediated setting to study the distribution of power in networks of negotiated exchange. The face-to-face setting used partitions to separate participants who were not connected to each other in the network, but subjects could hear one another's offers as well as extraneous comments. In the computer-mediated setting, subjects were seated in separate rooms and could not see or hear others; thus, influence from such comments was avoided. Although the data generally fit predictions in both settings, there were some indications that comments related to justice may have suppressed power use in some conditions.

More generally, computer-mediated interaction allows information and communication to be as highly constrained or as free as desirable, according to the research objectives. Earlier technologies were more restrictive. In early research on reciprocal exchange, for example, subjects interacted via computer-operated human test consoles that contained only push buttons, stimulus lights, and counters (Molm, 1981). In this setting, information was necessarily very minimal. When the test consoles were replaced with computer monitors, a much greater range of information was possible, and the kinds and amounts of information that subjects had about various aspects of the exchange—the size and shape of networks, the behavioral choices of other actors, and the value of benefits that others receive—could be varied or controlled.

Second, advances in computer networking capabilities have made the manipulation of exchange networks and subjects' assignments to positions in those networks a relatively simple matter. Subjects can be physically situated in a particular room and yet be assigned to any network, occupy any position within a network, and be connected (or not connected) to any other positions in the

network, as specified by the researcher. Computers have also made it possible to refine and control the bargaining process in negotiated exchange settings in ways that were not originally anticipated. Early experiments on negotiated exchanges simply allowed subjects to make offers and counteroffers to whomever they wanted, whenever they wanted, within a particular time frame. Later settings added more control over this process—for example, by specifying that all actors must make offers to all partners on each exchange opportunity, by limiting the number of rounds of offers and counteroffers that are possible, and by constraining the range of offers and counteroffers that can be made.

Third, computers have made the use of “programmed” or “simulated” partners easier, more sophisticated, and more realistic. Although many experiments on exchange are conducted on networks of all real subjects, certain research questions are best answered with the help of simulated partners whose behavior is either held constant or manipulated as one of the experimental factors. If, for example, the researcher is interested in comparing perceptions of justice—either distributive or procedural—in different exchange networks or different forms of exchange, it is typically desirable to manipulate or control the equality or inequality of exchange outcomes as well as some aspects of the partner’s behavior. More generally, any time the research focus shifts from interaction patterns at the level of relations or networks to the responses (behavioral or affective) of individual actors, it is desirable to control the behavior of the other actors in the network by using computer-simulated actors for those positions.

## B Web-Based Experiments

A recent development with potential implications for experimental work on exchange networks is the advent of Web-based experiments, which make use of the Internet to conduct experiments outside the laboratory. Rather than having subjects go to a laboratory to participate in a 2-hour experiment with other subjects (typically other students at the same university), participants from across the country—or from multiple countries—can participate in experiments in which they interact with others through the Internet. Web-based experiments potentially allow much larger and more complex exchange networks to be studied, using more diverse populations of subjects, with interaction extending over longer periods of time.

At the same time, the use of Web-based experiments raises a number of concerns, including the challenges of ensuring experimental control, effectiveness of manipulations, and security of data. It remains to be seen whether Web-based experiments will become widely used and, if so, what kinds of contributions they will make to the experimental study of exchange and exchange networks.

## V EXAMPLES OF EXCHANGE EXPERIMENTS

To illustrate how different researchers have tested theories of exchange and exchange networks in standardized laboratory settings, let us consider four

experiments. These particular experiments represent four distinct research programs and illustrate the diversity of work conducted in the exchange tradition. They employ different settings for studying different forms of exchange; place varying emphasis on structure or process, networks or relations; and study different exchange outcomes, including power and inequality, commitment, and trust. They also span more than 20 years of research in the development of the contemporary exchange tradition.

## A Cook and Emerson: Power and Equity in Exchange Networks

The first experiment is the classic [Cook and Emerson \(1978\)](#) study of how power and equity affect power use in exchange networks. The primary objective of this study was to test the central thesis of Emerson's power-dependence theory: that networks of power-balanced relations will produce equal benefits for actors, while networks of power-imbalanced relations will produce unequal benefits, with the distribution of benefits favoring the less dependent, more powerful actor. A second objective was to examine the constraining effects of equity concerns on the use of power, testing the prediction that power use would be lower under conditions in which significant equity concerns were expected to operate.<sup>9</sup>

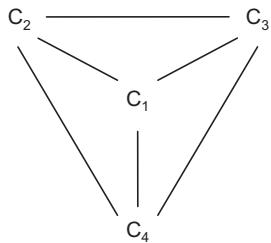
The researchers tested these predictions in the negotiated exchange setting described previously. Undergraduate student subjects were recruited on the basis of a desire to earn money and were randomly assigned to positions in one of two networks: a power-balanced network or a power-imbalanced network ([Figure 9.2a](#)). Relations in the networks were negatively connected, operationalized by allowing subjects to make an agreement with only one of their three alternative partners on each of 40 opportunities to exchange.

The balance or imbalance of power was manipulated by varying the *value* of actors' alternatives. In the power-balanced network, all potential exchange relations were of equal value (worth 24 points, divided through negotiations); thus, all actors in the network were in equivalent positions of power. In the power-imbalanced network, actors were *not* in equivalent positions of power. One actor, *A*, had three high-value (24 points) alternative exchange relations, whereas the three *Bs* had only one relation (with *A*) of high value and two others (with the other *Bs*) of low value (8 points). Thus, *A* had a power advantage over the *Bs*. The theory predicted that over time, power use would increase in *A*'s favor, as the *Bs* competed with each other to make the more valuable agreements with *A*, by offering *A* increasingly more of the 24 points.

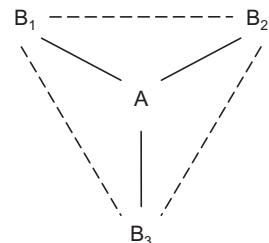
An important assumption of power-dependence theory that Cook and Emerson were testing was the principle that power use is *structurally determined*—in this case, by the availability of high-value exchange relations—regardless of actors' awareness of power or intent to use power. To test this central tenet, subjects

---

9. The study also examined the effects of emerging commitments among exchange partners on the use of power.

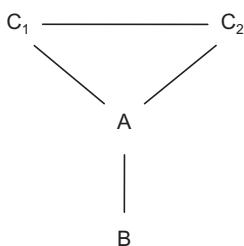


Power-Balanced Network

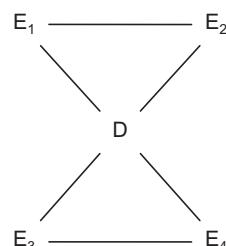


Power-Imbalanced Network

## (a) Networks Studied by Cook and Emerson (1978)

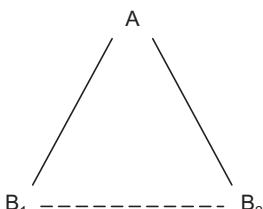


Stem Network

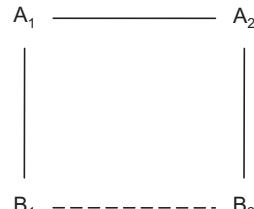


Kite Network

## (b) Weak-Power Networks Studied by Markovsky and Colleagues (1993)



High-Power Network



Low-Power Network

## (c) Networks Studied by Molm, Takahashi, and Peterson (2000)

**FIGURE 9.2 Networks in three exchange experiments.** Solid lines indicate relations with high exchange value; dashed lines indicate relations with low exchange value. (a) Networks studied by Cook and Emerson (1978); (b) weak-power networks studied by Markovsky and colleagues (1993); (c) networks studied by Molm and colleagues (2000).

were told only of their immediate connections to others in the network; they were unaware of the size or shape of the network as a whole.

The researchers manipulated equity concerns by varying the information subjects had about others' earnings. For the first 20 trials, subjects knew only

their own cumulative earnings; then, they were told the cumulative earnings of *all* participants in the network. That meant that, for the last 20 trials, subjects in the power-imbalanced networks bargained with knowledge of the inequalities in benefits between actors in the *A* and *B* positions. This knowledge should trigger equity concerns, which were predicted to reduce power use in the power-imbalanced networks.

The design of this study, then, was a mixed design, containing a between-subjects variable (power balance/imbalance) and a within-subjects variable (equity). In addition, four trial blocks of five trials each were nested within the two conditions of the equity manipulation, allowing the examination of change over time. The primary dependent variable was power use, measured as a difference score (the difference in benefits received by the two actors in an agreement). Because of the interdependencies between subjects in a network and transactions across the exchange period, the entire four-person network, interacting through 40 3-minute transaction periods, was the statistical unit for the analyses. Fourteen networks were studied in each of the two power conditions.

Findings supported the main predictions: mean power use was significantly greater in the power-imbalanced network than in the power-balanced network, and it increased over time up to the time of the equity manipulation. Then, as predicted, knowledge of everyone's earnings produced a significant and marked decrease in power use in the power-imbalanced networks.

## B Markovsky and Colleagues: Strong and Weak Power

Theoretical advances sometimes occur when experiments produce unexpected findings that challenge current theoretical formulations and lead to modifications of a theory and further experimental test. This study, conducted by Markovsky, Skvoretz, Willer, Lovaglia, and Erger in 1993, is an example of this process.

In the original formulation of network exchange theory, power in exclusionary networks (called negatively connected in power-dependence theory) was measured solely by the GPI. This index predicted power use quite accurately in some networks, but in others, weak power differences emerged even though the GPI predicted no differences in power. In this study, Markovsky and colleagues developed the distinction between "strong power" networks and "weak power" networks and tested a modified version of the theory designed to do a better job of predicting the small but consistent differences that occur in weak power networks.

Powerful actors in strong- and weak-power networks vary in their capacity to exclude others consistently from negotiated exchanges without suffering costs themselves. In strong power networks, one or more disadvantaged actors are excluded on every opportunity from exchanges with a powerful actor, at no cost to the powerful actor. For example,  $B_1-A-B_2$  is a strong power network because one of the *Bs* is always excluded from exchange with *A* on each

opportunity. According to the path-counting algorithm of the GPI, this occurs because  $A$ 's two alternatives—the two  $B$ s—lie on paths of length 1 and are highly available for  $A$ , whereas the  $B$ s' single alternative— $A$ —lies on a path of length 2 and is less available to either  $B$ .

In weak power networks, either all positions are equally subject to exclusion or no position can consistently exclude another without incurring cost to itself. For example,  $B_1-A_1-A_2-B_2$  is a weak power network:  $A_1$ 's immediate alternative to  $B_1$ —the other  $A$ —lies on a path of length 2 and has another potential partner. The two  $A$ s can exclude the  $B$ s only by exchanging with each other, but this action is costly because agreements between the equal-power  $A$ s should provide each with no more than half the benefits. Nevertheless, positions in weak-power networks still differ in their *probabilities* of inclusion or exclusion, and an iterative refinement of the GPI was developed to capture these more subtle differences in power. Markovsky and colleagues then tested the predictions of the modified theory in two weak-power network structures, a four-actor network called the “stem” and a five-actor network called the “kite” (see [Figure 9.2b](#)).

The negotiated laboratory setting included the features described previously: subjects' computer screens displayed full information of the network configuration, current offers, and completed exchanges, and subjects were rotated (by a software configuration) through the different network positions, with four negotiation rounds in each position. A total of eight networks were run in the stem configuration and six networks in the kite configuration. Predictions specified the relative ordering of exchange outcomes (i.e., which actor of each two-party pair would get more from his or her agreements) for structurally distinct relations within the two networks. The units of analysis were the observed exchanges and nonexchanges among pairs of subjects, and a dummy-variable, constrained-regression analysis was used to estimate positional effects ([Skvoretz & Willer, 1991](#)). The findings confirmed the predicted weak-power effects: in the stem network, agreements between  $A$  and either  $B$  or  $C$  favored  $A$ ; in the kite network, agreements between  $D$  and  $E$  favored  $D$ .<sup>10</sup>

## C Lawler and Yoon: Commitment in Exchange Relations

Lawler and Yoon also investigated the effects of power structures in a series of experiments testing relational cohesion theory, but the exchange outcome in which they were interested was the development of cohesion and commitment between exchange partners, not power use, and their focus was on dyadic relations, not exchange networks. In a 1996 experiment, they tested a model predicting that Emerson's two dimensions of structural power (total power-dependence and relative power-dependence—or power imbalance—in the relation) affect the frequency of agreements between two parties, and that frequent,

---

10. This study also tested hypotheses about the power use of “experienced” versus “inexperienced” subjects and about strong versus weak power networks (by comparison with earlier experiments).

successful exchanges produce positive emotions that lead to perceptions of the relation as a cohesive unit and various forms of commitment behaviors. They conducted three experiments, each examining a different kind of commitment behavior: giving token gifts to the partner; staying with the partner in the face of alternatives; and contributing to a joint venture. The gift-giving experiment is described here.

Undergraduate subjects were assigned to roles representing either Alpha Company or Beta Company. Alpha was trying to buy iron ore from Beta, and the task of both representatives was to negotiate the best possible deal for their companies. Subjects knew only their own profit, not their negotiation partner's profit, from any agreements reached. The subjects negotiated over 12 episodes but were told to expect 15, a procedure designed to prevent "end effects" (i.e., strategies based on the knowledge that the partner cannot respond contingently after the last episode). In each episode, if Alpha and Beta failed to reach an agreement, their profit was determined by an agreement with a hypothetical alternative supplier or buyer. Subjects knew the probabilities of different profits that they and the other actor could earn from the alternative; a random drawing determined the actual profit received.

The two dimensions of power were crossed in a factorial design (with 20 dyads per cell) and manipulated by varying the probabilities of agreements with the hypothetical alternative. In the equal-power conditions, the expected value from the alternative was 60 points for both actors when total power was high and 80 points for both actors when total power was low (i.e., low total power implied that dyad members were less dependent on each other relative to the alternative). In the unequal-power conditions, the expected value of the alternative was greater for the high-power actor than for the low-power actor: 75/50 under low total power and 100/65 under high total power. Thus, although the dyad was not embedded in an exchange network in these experiments, the same notion of alternatives underlying power and dependence relations governs the manipulation of power.

All of the other variables in the analysis were measured: the proportion of negotiation episodes in which Alpha and Beta reached agreement; the emotions the subjects reported experiencing (measured after episodes 4 and 8, on dimensions of pleasure/satisfaction and interest/excitement); perceptions of the cohesiveness of the relationship (measured after episode 8); and the gift-giving measure of commitment. Giving a token gift—a voucher that subjects could exchange for pieces of candy after the experiment—was an option after each of episodes 9 through 12. Subjects could give the voucher to their partner or keep it for themselves; they were told they would not know if the other gave them gifts until the experiment was over.

The analysis tested the endogenous process predicted by the theory: that high total power and equal power increase exchange frequency, which increases positive emotions, which increase perceptions of relational cohesion, which increase the frequency of gift-giving as a behavioral indicator of commitment.

A series of regression analyses, testing the predicted indirect paths of the model, generally supported the theory.

[Lawler et al. \(2000\)](#) used a variant of this setting to extend the theory of relational cohesion to productive exchange. Rather than dyads negotiating two-party agreements, triads negotiated three-party agreements. Again, relative dependence (equal vs. unequal) and total dependence (high vs. low) were manipulated through the use of a hypothetical alternative, and effects on the endogenous processes were examined. Comparisons with the 1996 dyadic experiment showed that agreements were less frequently reached in the triad than in the dyad. Consequently, cohesion and commitment were more problematic: positive emotions were lower; perceptions of perceived cohesion were lower; and fewer gift opportunities were used.

## D Molm, Takahashi, and Peterson: Risk and Trust

So far, all the examples we have discussed have been experiments on negotiated exchange. In a series of experiments, Molm and colleagues compared negotiated exchange with reciprocal exchange, testing the thesis that reciprocal exchange produces stronger trust, affective regard, and relational solidarity than negotiated exchange, and developing a theory of reciprocity in exchange. That theory proposes that variations in the structure of reciprocity in different forms of exchange produce these effects through three intervening mechanisms: the structural risk of nonreciprocity; the expressive value conveyed by acts of reciprocity; and the relative salience of the cooperative or conflictual elements inherent in the mixed-motive structure of exchange ([Molm, 2010](#); [Molm et al., 2007](#)).

The first experiment in this program tested the classical prediction that trust is more likely to develop between exchange partners when exchange occurs without explicit negotiations or formal agreements. Negotiated exchanges with binding agreements provide assurance against exploitation, but they give actors little opportunity to develop trust in one another. Acts of trust and attributions of trustworthiness can only be made in situations that involve some risk and uncertainty; that is, the partner must have both the incentive and the opportunity to exploit the actor. Reciprocal exchanges provide the necessary risk and uncertainty for trust to develop; actors in these exchanges initiate exchange without knowing whether or when the partner will reciprocate. If the partner behaves in a trustworthy manner under these conditions, then the actor's trust in the partner should increase.

[Molm et al. \(2000\)](#) tested this logic in a laboratory experiment that compared negotiated and reciprocal forms of exchange in equivalent, negatively connected networks of three or four actors (see [Figure 9.2c](#)). Within each network, a power advantage for A over B was created by giving A a high-value (16 points) alternative to exchange with B, whereas B had only a low-value (4 points) alternative to exchange with A. Across networks, the strength of A's power advantage was manipulated by varying the availability of A's alternatives. Note that these

manipulations combine aspects of both the Cook and Emerson manipulation of power (based on the value of alternatives) and the Markovsky and associates manipulation of power (based on availability of alternatives). A factorial design crossed the form of exchange (negotiated or reciprocal) with the power advantage for the powerful actor (high or low power). Ten networks were studied in each of the four conditions.

In the negotiated exchange setting, subjects jointly negotiated the division of a fixed amount of benefit on each of a series of exchange opportunities; on each opportunity, subjects had five rounds of offers and counteroffers to make an agreement with another subject to whom they were connected in the network. In the reciprocal exchange setting, as described previously, each actor gave points to one of his or her partners on each opportunity, without knowing whether the partner would reciprocate. The total value of exchange within relations was held constant across the two forms of exchange. As in the [Cook and Emerson \(1978\)](#) and [Lawler and Yoon \(1996\)](#) experiments, subjects did not know the value their partners received from their exchanges or their partners' cumulative earnings. Their knowledge of the network structure was also limited to their immediate exchange relations. Subjects' trust in their partners, as well as other integrative outcomes (positive affect and feelings of commitment), was measured at the end of the exchange period by a series of semantic differential scales.

The manipulation of the form of exchange affected the *risk* predicted to be necessary for the development of trust (reciprocal exchanges are riskier than negotiated exchanges); the manipulation of power imbalance affected two behaviors used as indicators of the partner's *trustworthiness*: behavioral commitment and equality of exchange, both of which were greater when power imbalance was low. Molm and colleagues predicted that average trust would be greater in the reciprocal exchange relations than in the negotiated exchange relations, that an actor's trust in the partner would increase with the partner's behavioral commitment to the actor and with the equality of their exchange, and that the effects of these behaviors on trust would be stronger in the reciprocal exchange conditions than in the negotiated exchange traditions.

These predictions were supported. Subsequent experiments in the research program demonstrate that these effects are quite broad; reciprocal exchanges produce more positive outcomes than negotiated exchanges on a wide range of relational measures—not only greater trust, affective regard, and feelings of commitment but also stronger perceptions of fairness and of relational solidarity or social unity (see [Molm, 2010](#)).

[Molm et al. \(2007\)](#) tested an extension of this logic to generalized exchange. In generalized exchange, reciprocity is not only uncertain but also *indirect*; that is, A's giving to B is reciprocated not by B but by another actor in the generalized exchange system. Thus, the risk of nonreciprocity is even greater. They compared chain-generalized exchange networks of three or four actors with equivalent networks of negotiated and reciprocal exchange; in this study, power was balanced in all networks. As predicted, trust was greater in

the chain-generalized exchange with indirect reciprocity than in either of the forms of exchange with direct reciprocity.

## VI ASSESSMENT AND FUTURE PROSPECTS

The experimental tradition of social exchange research is one of the best examples in sociology of sustained theory construction and testing through programmatic, cumulative research. The sheer quantity of research during the past 35 years stands in sharp contrast to the earlier years of exchange theory in sociology, when little empirical work was conducted. The combined efforts of many researchers, conducting long-term programs of research testing different exchange theories, have produced a strong empirical base that offers substantial support for the social exchange perspective.

The use of standardized laboratory settings has contributed to the cumulation of knowledge in the field by making it easier for researchers to compare results across experiments. The process of theory construction necessarily involves numerous experimental tests, with each experiment investigating only one part of a theory. Most of the research programs reviewed in this chapter also involve both deduction and induction, with experiments playing a vital role in both aspects of theory construction. Work typically begins with experiments testing theoretically derived hypotheses (deduction); results of those experiments sometimes challenge theoretical assumptions and pose empirical puzzles that must be solved theoretically (induction), and new experiments then test the revised theory (deduction).

The study of power in exchange networks dominated the field during much of this period and engaged the attention of numerous researchers working in several distinct programs. This work greatly increased our knowledge of how particular network structures affect patterns of exchange, distributions of power, and inequality. More recently, other researchers have applied the same kind of sustained effort to the development and testing of theories addressing more integrative aspects of exchange, such as the development of trust, affective bonds, and commitment, and incorporating the role of affect and emotions in exchange relations.

Promising trends in current research suggest that, in future decades, we will see significant progress on two important topics that have been relatively neglected until recently. The first is the study of change—changes in relationship dynamics and changes in network structures. Despite the ubiquity of change in social life, virtually all experimental studies of social exchange have examined development of relationships within static structures and without change in other objective features of the relation. Recently, however, researchers have begun to ask how different histories of negotiated and reciprocal exchange, experienced either sequentially or concurrently, affect relationships (Cheshire, Gerbasi, & Cook, 2010; Molm, Whitham, & Melamed, 2012). This work has the potential to build bridges between the experimental tradition of exchange and

research in natural settings conducted by sociologists in other fields, particularly organizational scholars.

Research on structural change remains rare, however. As the examples in this chapter illustrate, researchers commonly assign subjects to positions in particular network structures and then examine the effects of those structures on behavioral and affective outcomes. Networks are rarely static, however; they expand and contract, network connections change, and the value of resources attached to different positions varies over time. Studying what produces structural change, how change affects established patterns of interaction, and how the structural history of a network alters its current impact are all topics that could and should be pursued in laboratory experiments on network change. Renewed interest in coalition formation, as a form of structural change, is a promising step in this direction (e.g., [Simpson & Macy, 2004](#)).

Second, as this review suggests, researchers are beginning to direct much-needed attention to more collective forms of exchange, including productive exchange and various forms of generalized exchange. The sociological tradition of exchange that George Homans initiated explicitly rejected the study of generalized (indirect) exchange in favor of a focus on direct exchange. But both generalized and productive forms of exchange are highly relevant to contemporary concerns with the development of trust and solidarity in social life, and both are pervasive throughout society. By combining laboratory experiments on relatively small collective systems with computer-simulated studies of larger systems, researchers can study the role of these more collective forms of exchange in creating strong bonds in groups and networks. Such work promises to reconnect the contemporary exchange tradition in sociology with its early anthropological roots, thus bringing the tradition full circle.

## REFERENCES

- Bacharach, S. B., & Lawler, E. J. (1981). *Bargaining: Power, tactics, and outcomes*. San Francisco: Jossey-Bass.
- Bienstock, E. J., & Bonacich, P. (1992). The core as a solution to exclusionary networks. *Social Networks*, 14, 231–243.
- Blau, P. M. (1964). *Exchange and power in social life*. New York: Wiley.
- Cheshire, C., Gerbasi, A., & Cook, K. S. (2010). Trust and transitions in modes of exchange. *Social Psychology Quarterly*, 73, 176–195.
- Cook, K. S., & Emerson, R. M. (1978). Power, equity and commitment in exchange networks. *American Sociological Review*, 43, 721–739.
- Cook, K. S., Emerson, R. M., Gillmore, M. R., & Yamagishi, T. (1983). The distribution of power in exchange networks: Theory and experimental results. *American Journal of Sociology*, 89, 275–305.
- Cook, K. S., & Gillmore, M. R. (1984). Power, dependence, and coalitions. In E. J. Lawler (Ed.), *Advances in group processes: Vol. 1*. (pp. 27–58). Greenwich, CT: JAI Press.
- Emerson, R. M. (1962). Power-dependence relations. *American Sociological Review*, 27, 31–41.

- Emerson, R. M. (1972a). Exchange theory, part I: A psychological basis for social exchange. In J. Berger, M. Zelditch, Jr., & B. Anderson (Eds.), *Sociological theories in progress: Vol. 2*. (pp. 38–57). Boston: Houghton–Mifflin.
- Emerson, R. M. (1972b). Exchange theory, part II: Exchange relations and networks. In J. Berger, M. Zelditch, Jr., & B. Anderson (Eds.), *Sociological theories in progress: Vol. 2*. (pp. 58–87). Boston: Houghton–Mifflin.
- Friedkin, N. E. (1992). An expected value model of social power: Predictions for selected exchange networks. *Social Networks*, 14, 213–229.
- Homans, G. C. (1961). *Social behavior: Its elementary forms*. New York: Harcourt Brace and World.
- Kollock, P. (1994). The emergence of exchange structures: An experimental study of uncertainty, commitment, and trust. *American Journal of Sociology*, 100, 313–345.
- Lawler, E. J. (1992). Power processes in bargaining. *The Sociological Quarterly*, 33, 17–34.
- Lawler, E. J. (2001). An affect theory of social exchange. *American Journal of Sociology*, 107, 321–352.
- Lawler, E. J., Thye, S. R., & Yoon, J. (2000). Emotion and group cohesion in productive exchange. *American Journal of Sociology*, 106, 616–657.
- Lawler, E. J., Thye, S. R., & Yoon, J. (2008). Social exchange and micro social order. *American Sociological Review*, 73, 519–542.
- Lawler, E. J., Thye, S. R., & Yoon, J. (2009). *Social commitments in a depersonalized world*. New York: Russell Sage Foundation.
- Lawler, E. J., & Yoon, J. (1996). Commitment in exchange relations: Test of a theory of relational cohesion. *American Sociological Review*, 61, 89–108.
- Markovsky, B., Skvoretz, J., Willer, D., Lovaglia, M., & Erger, J. (1993). The seeds of weak power: An extension of network exchange theory. *American Sociological Review*, 58, 197–209.
- Markovsky, B., Willer, D., & Patton, T. (1988). Power relations in exchange networks. *American Sociological Review*, 53, 220–236.
- Molm, L. D. (1981). The conversion of power imbalance to power use. *Social Psychology Quarterly*, 16, 153–166.
- Molm, L. D. (1997). *Coercive power in social exchange*. Cambridge, UK: Cambridge University Press.
- Molm, L. D. (2010). The structure of reciprocity. *Social Psychology Quarterly*, 73, 119–131.
- Molm, L. D., Collett, J. L., & Schaefer, D. R. (2006). Conflict and fairness in social exchange. *Social Forces*, 84, 2331–2352.
- Molm, L. D., Collett, J. L., & Schaefer, D. R. (2007). Building solidarity through generalized exchange: A theory of reciprocity. *American Journal of Sociology*, 113, 205–242.
- Molm, L. D., Takahashi, N., & Peterson, G. (2000). Risk and trust in social exchange: An experimental test of a classical proposition. *American Journal of Sociology*, 105, 1396–1427.
- Molm, L. D., Takahashi, N., & Peterson, G. (2003). In the eye of the beholder: Procedural justice in social exchange. *American Sociological Review*, 68, 128–152.
- Molm, L. D., Whitham, M. M., & Melamed, D. (2012). Forms of exchange and integrative bonds: Effects of history and embeddedness. *American Sociological Review*, 77, 141–165.
- Schaefer, D. R., & Kornienko, O. (2009). Building cohesion in positively connected exchange networks. *Social Psychology Quarterly*, 72, 384–402.
- Simpson, B., & Macy, M. W. (2004). Power, identity, and collective action in social exchange. *Social Forces*, 82, 1373–1409.
- Skvoretz, J., & Willer, D. (1991). Power in exchange networks: Setting and structural variations. *Social Psychology Quarterly*, 54, 224–238.

- Stolte, J. R., & Emerson, R. M. (1977). Structural inequality: Position and power in network structures. In R. L. Hamblin, & J. Kunkel (Eds.), *Behavioral theory in sociology* (pp. 117–138). New Brunswick, NJ: Transaction.
- Szmatka, J., & Willer, D. (1995). Exclusion, inclusion and compound connection in exchange networks. *Social Psychology Quarterly*, 55, 123–131.
- Thibaut, J. W., & Kelley, H. H. (1959). *The social psychology of groups*. New York: Wiley.
- Thye, S. R., Lovaglia, M. J., & Markovsky, B. (1997). Responses to social exchange and social exclusion in networks. *Social Forces*, 75, 1031–1047.
- Yamagishi, T., & Cook, K. S. (1993). Generalized exchange and social dilemmas. *Social Psychology Quarterly*, 56, 235–248.
- Yamagishi, T., Cook, K. S., & Watabe, M. (1998). Uncertainty, trust and commitment formation in the United States and Japan. *American Journal of Sociology*, 104, 165–194.
- Yamagishi, T., Gillmore, M. R., & Cook, K. S. (1988). Network connections and the distribution of power in exchange networks. *American Journal of Sociology*, 93, 833–851.

## Chapter 10

# Social Dilemma Experiments in Sociology, Psychology, Political Science, and Economics

Jane Sell

Bruce Reese

*Texas A&M University, College Station, Texas*

### I INTRODUCTION

Social dilemmas occur in settings in which there is a conflict between individual short-term incentives and overall group incentives (see Dawes, 1980; Kollock, 1998). There are two types of social dilemmas: public goods in which people must decide whether or not to contribute; and resource goods in which people must decide whether to consume or refrain from consuming. Public goods are things that are available for all group members to use; examples include public radio and national defense. Resource goods are maintained for group members' use at some future time; examples include the maintenance of fragile rain forests and the protection of endangered species.

Public and resource goods are quite different from private goods. For private goods, an individual pays for a good (and thus reveals her preference) and then can consume the good individually. With resource and public goods, there is no direct relationship between contribution (or restraint from taking) and private acquisition. So, for example, regardless of whether you contribute to public radio, you can still consume or listen to public radio. This property, called non-excludability, is at the heart of the incentive issue of social dilemmas. A person can refuse to contribute but still consume the good; that is, the person can "free-ride" on the contributions of others. As this is the case, from the point of view of the individual, it is tempting to not contribute and to hope that others will contribute. But if all share that point of view, nobody contributes and the public or resource good is not provided. This is the dilemma.

We are faced with these dilemmas everywhere. Sometimes these dilemmas involve our relationships with intimates—for example, although everyone in a

family might prefer it if everyone else cleaned, if everyone waited for others, the house would never be clean. Sometimes these dilemmas involve institutions or countries. On an international level, if some countries voluntarily reduce their pollution rates, other countries benefit, even if they themselves do not reduce rates.

Because social dilemmas are so pervasive, they have been the subject of attention for sociology, psychology, political science, and economics. There is a different emphasis in these different fields, and there are many examples of interdisciplinary research. The interdisciplinarity has been aided by the acceptance and use of experimental research. We examine some central issues in the study of social dilemmas that cut across disciplines. In particular, we emphasize structural aspects of social dilemmas.

We first consider the basic structure of the social dilemma. For more than 25 years, there has been a movement toward more precision in specifying the type of social dilemma. For the most part, this has been in response to unexpected, sometimes surprising results from both laboratory and field experiments.

## II DIFFERENT KINDS OF SOCIAL DILEMMAS

### A Two-Person Dilemmas

There is a long tradition of two-person social dilemmas. In fact, the most famous of the two-person dilemmas is termed the prisoner's dilemma or PD (named for a story about two prisoners, their separation, and the manner in which police try to structure the incentive for each to inform on the other). These two-person dilemmas are often characterized by a  $2 \times 2$  table typology in which two choices for each person are characterized: cooperate or defect ([Figure 10.1](#)). The payoffs associated with cooperating with or defecting from an other (or partner) determine the incentive structure. Listing Person 1 first and Person 2 second, the four possible options then are cooperate/cooperate, cooperate/defect, defect/cooperate, and defect/defect. If the payoffs associated with cooperate/cooperate are the highest, then the incentive dilemma (the defining characteristic of social

		Actor 1	
		Cooperate	Defect
Actor 2	Cooperate		
	Defect		

**FIGURE 10.1** Illustration of choices associated with two actors in a prisoner's dilemma.

dilemmas) is absent, but there may still be a coordination problem. That is, how do the participants make sure that everyone knows what is the best choice?

The two-person dilemma has structured an enormous amount of research. Much of this research has centered on the issue of trust (discussed later in this chapter). These kinds of scenarios, although tremendously important for two-party interactions, are considered apart from larger group analysis of social dilemmas. There are some sound theoretical reasons for a division based on size of the group. As [Dawes \(1980\)](#) indicated, when only two people are involved, the source of the defection or cooperation is completely known by both parties when they know the outcome. Each person of course knows his own behavior, and when he learns his outcome, he can tell immediately whether the partner cooperated or defected. Because this is the case, monitoring is complete, and sanctions can be specifically targeted to a single person. In cases in which multiple actors are involved (usually termed N-PDs), responsibility for actions is more diffuse and consequently monitoring is more problematic. [Kollock and Smith \(1996\)](#) correctly note that it is not necessarily the case that two-person and  $N$ -person dilemmas are completely different. Thus, at times,  $N$ -person dilemmas could have the same characteristics as the two-person case.

[Kollock and Smith's \(1996\)](#) observation is important and relates to the direct tie between the theoretical questions being asked and the development of an experimental test of those questions. So, for example, designing the experiment is a matter of what theoretical question is being asked—not necessarily the same thing as the particular number of participants. It is possible to design an experiment with many participants that involves the complete monitoring that is the default case for a two-person dilemma.

The intimate link between the theoretical question asked and the design is, of course, the major strength of the experiment. Because social dilemmas are so pervasive in everyday situations, experience based on actual settings has sometimes been taken for granted. The power of the experimental method to separate out factors has been instrumental sometimes in demonstrating that investigators were wrong about what factors lead to what behaviors.

## B N-Person Social Dilemmas

As mentioned previously, social dilemmas can be categorized into two groups: public goods and resource goods. Each of these has a classic statement or set of statements. For public goods, there is the theoretical explication by [Paul Samuelson \(1954, 1955\)](#) and [Mancur Olson's \(1965\) \*The Logic of Collective Action\*](#). The goods and collective movements described in these pieces survive through contributions of members. Garrett [Hardin's \(1968\)](#) “The Tragedy of the Commons” is certainly the most mentioned article depicting resource goods and the temptation to take from the public domain. The tragedy mentioned by Hardin was the temptation for herders to overgraze their livestock on the

commons and cause the degradation of the very resource that was necessary for themselves as well as the community.

For the most part, these two kinds of goods were viewed as equivalent, and so the theoretical principles developed in one were generally applied to the other. Such equivalence seemed reasonable, especially from an economics standpoint because the payoffs created the same set of incentives (Ledyard, 1995). In all cases, the social dilemma incentives favored individual-level defection (or noncooperation) at each individual point in time. However, some in psychology argued that the two types of settings were different from a social psychological point of view (see Brewer & Kramer, 1986). Sell (1988) argued that the distinction between them could be gauged on the basis of the setting and then actors' responses to those structures. In particular, she argued that whereas public goods could be characterized by the production function, which linked the contributions or resources of group members to the public good, resource goods (often created experimentally to "mirror" natural resources) are characterized by a replenishment function that creates a different type of good.

Brewer and Kramer were the first to investigate the difference in 1986; this was followed by several other studies (see for example McCusker & Carnevale, 1995; Messick, Allison, & Samuleson, 1988; Rutte, Wilke, & Messick, 1987; Schwartz-Shea & Simmons, 1995). Results were mixed. Although all this research was experimental, there were differences in the good being examined. Son and Sell (1995) and Sell and Son (1997) argued that some of the differences might be due to the nature of the public good, but that the static versus dynamic nature of the setting also could dramatically affect outcomes, partly in terms of uncertainty of outcome. Theoretically, there is a major difference between a one-time event or decision and a multiple series of decisions, and also a difference between a static series in which each decision is like every other and a series of decisions in which a replenishment function transforms the resource.

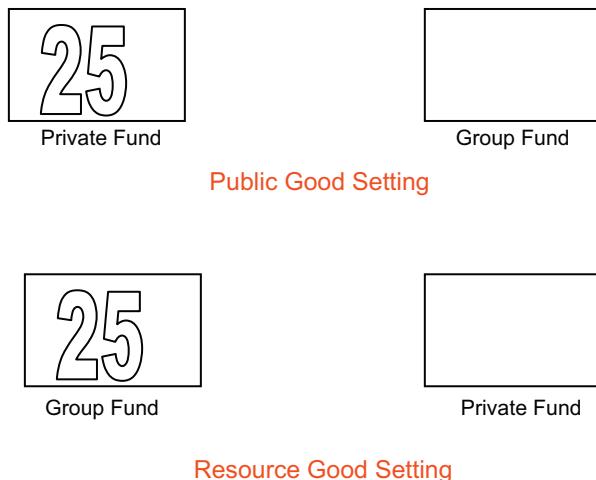
In a series of studies designed to determine how the type of good affected levels of cooperation (either restraint from taking or contributing), Sell and Son (1997) and Sell, Chen, Hunter-Homes, and Johansson (2002) tried to directly compare giving versus restraint in exactly the same settings. One-time decision-making was compared to one-time decision-making; repeated decision-making was compared to repeated decision-making; and dynamic decision-making was compared to dynamic decision-making. To make comparisons, the simplest version of the public good setting and the resource good setting was utilized. The experimental scenario was adapted from the standard linear public goods setting that was developed and modified by Gerald Marwell and Ruth Ames in their research beginning in 1979 (Marwell & Ames, 1979), and then Mark Isaac, James Walker, and Susan Thomas in 1984 (Isaac et al., 1984). For public goods, subjects are provided with an initial *endowment*, an amount of money individually allocated to the subjects. This endowment is usually specified in terms of tokens so that the tokens can take on different monetary values. The subject then is given the choice to keep his or her endowment in the private fund or invest in a group

fund, a fund that has a different payoff than the private fund and is shared with group members regardless of whether those members have contributed or not. This standard linear public goods setting has a number of advantages. It is simple, and it provides a simple way to quantify the incentive structure. The different values for the different funds allow a method of calculating what [Isaac and Walker \(1988\)](#) termed the marginal per capita return (MPCR) on the public good. Thus, for example, in a particular experiment, if there are four members of a group and if every token in that experiment is worth 1¢ in the private fund, but worth 3¢ in the public fund (the public good), the MPCR on the public good is 0.75. To ensure the defining property of a public good, the MPCR on the public good (at a given point in time) is always less than 1. Such measurement enables comparison across different settings and more precise estimates of the nature of the incentives.

In comparing public goods and resource goods, it was important that the framing or language aspects were kept minimal because language can affect subjects' decisions. So, for example, [Andreoni \(1995\)](#) found that framing choices as focusing on others' benefits or losses rather than just making decisions created increased contributions to the public good. Consequently, no descriptions of actual settings were used, and wording differences between the resource and public goods were minimized. In computerized versions of the experiments, two simple blue boxes appeared on a subject's screen. In the public good setting, an individual's tokens appeared in the blue box marked "private fund," and subjects could choose whether to keep tokens in the private fund, which paid a certain amount and was not shared with others. Their other choice was to contribute to a group fund that paid another amount and was divided (equally) among all group members. In the resource good setting, the same blue boxes appeared, but the tokens appeared first in the group fund and subjects could choose whether to keep the tokens in the group fund (which again was divided equally among group members) or withdraw tokens to invest in the private fund ([Figure 10.2](#)). For both these cases, the payoffs were exactly the same for the two funds, and these payoffs instantiated the property of the social dilemma such that for any given period of time, the individual's payoff for the private fund was more than his or her payoff from the group fund.

When the experiments were not computerized (when, for example, this was not possible in cultures where computers were not available), the blue boxes were replaced by two different columns. Those subjects randomly assigned to the resource good condition found their token amount listed in the group fund column initially, whereas those randomly assigned to the public good condition found their token amount listed in the individual fund initially.

Record sheets indicated other aspects of the good setting, such as replenishment rates and results from previous trials. In this manner, each subject could see what behaviors had occurred in the past, and these were made salient. (This was particularly important because some studies had been criticized for being confusing or unclear about the payoffs and the information.)



**FIGURE 10.2** Comparison of screens for static public goods and resource goods.

Results from these studies indicated that, indeed, resource and public goods did generate different levels of cooperation for static settings. In particular, in violation of traditional expected utility formulations, but explained through alternative theoretical perspectives, resource goods settings generated more cooperation than did public goods settings. When tokens started out in the group fund, subjects were more likely to cooperate than when the tokens started out in their own individual fund. Furthermore, this process builds up over time; cooperation builds more cooperation.

Although these results are important, there is a central mystery in them as in many results: why do people cooperate *at all* in any setting that creates a social dilemma? The incentive structure seems to dictate that people should *not* cooperate because it is not in their best interest. Much research was directed at answering this question, and both theoretical and empirical advancement resulted.

### III REJECTION OF THE STRONG FREE-RIDING HYPOTHESIS

From the viewpoint of traditional economic models, public goods must entail coercion to sustain them. This is the case because if an actor is rational and seeking only to maximize her own utility, she will not contribute. Furthermore, with such goods, there must be some way to develop measures of willingness to pay (or preferences) to optimize the welfare of those in the particular setting. This is problematic because unlike the private good situation in which there is no incentive to misrepresent preferences, there *is* incentive for people to understate preferences because the good can be consumed by all. This theoretical perspective was explicated by Paul Samuelson in 1954, 1955, and 1958 (Samuelson, 1954, 1955, 1958).

Much of welfare economics takes these two issues as a given, but experimentalists sought to explore this, initially to develop an important baseline.

However, there were problems in the empirical analyses of such settings. [Bohm \(1972\)](#) designed a survey to study individuals' willingness to pay for a new television show. One group was told that their personal cost would be tied to their stated willingness to pay while another group was told that their cost was irrespective of their stated willingness to pay. The primary prediction was that those who were asked their true willingness to pay would underestimate their willingness (and therefore would free-ride), whereas those whose cost was independent of their statements would overstate preferences to ensure that the show would be provided. However, the predictions were not supported. There was no difference in the willingness to pay among groups; the free-rider hypothesis was not supported.

This prompted a series of further tests. Many contended that there were interpretation problems with the Bohm experiment. Perhaps subjects were confused about the good; preferences were not induced and not known. Other researchers conducted studies to try to eliminate different alternative interpretations for the results.

Sociologists Gerald Marwell and Ruth Ames were interested in these general phenomena as well and conducted a series of experiments in natural settings. In a 1979 study, the researchers contacted a fairly large number of high school students (256). Students were contacted several times by phone and also received directions by mail. They were told that they could invest in two types of exchanges. The individual exchange was a sure investment so that for every token invested, the subject received a set amount. The group exchange, however, was shared by all group members irrespective of their individual contributions. This group exchange constituted the public good and contained a provision point—a point beyond which the public good payoff was greatly increased. As predicted by economic theory, if there was an individual in the group whose interest exceeded the cost of its provision, that group invested far more in the public good than did other types of groups.

However, other effects predicted by basic economic theory were not supported. Important among those predictions, far less free-riding occurred than would be predicted. Following up the 1979 research, and responding especially to critics who claimed that their experimental design could not rule out effects that might result from misunderstanding, [Marwell and Ames \(1980, 1981\)](#) systematically replicated parts of their 1979 research to eliminate several plausible alternative reasons for their results. For instance, to address critics who claimed that the stakes were not high enough, in one study they increased the stakes fivefold and found that it did not affect subjects' behavior.

In essence, these first groups of studies demonstrated that the strong free-riding hypotheses did not hold. This is an interesting and important contribution of experimental methodology because the control possible through the experimental design in terms of separation of effects of reputation, anticipation of future interaction, and knowledge of the characteristics of others in the group could be held constant. Even when possible interpretations, such as participants

somehow trying to please the experimenters, were controlled for, contributions never completely disappeared. Although the rate of free-riding could be increased by making it more materially lucrative, it never became universal. For sociologists and psychologists, the reason seems clear: there was a normative aspect to the phenomenon even when all usual normative trappings and language were stripped away. Economists have given this residual level of cooperation various terms, including “warm glow” (Andreoni, 1995). Some economists have argued that there may be a biological basis for cooperation that may need to be incorporated into economic theories of exchange. Kevin McCabe and Vernon Smith (2005), for example, argue that humans engage in a “goodwill accounting” that is part of humans’ evolutionary heritage. In addition, there has been a flurry of theoretical and empirical activity regarding evolutionary models that demonstrate the feasibility of many different types of cooperative strategies (Bowles & Gintis, 2004; Nowak, 2006).

## IV TESTING PAYOFF PROPERTIES

Although people may not completely free-ride, it is still common, and in the experiments it displayed predictable patterns according to the incentive structure; for example, the MPCR on a public good does make a difference. When the MPCR is lower, contributions are also lower (Isaac & Walker, 1988). Using the MPCR also allows the development of precise comparisons for cases in which some players receive high returns while others receive low returns (used, for example, by Fisher, Isaac, Schatzberg, & Walker, 1995).

This scenario can be modified in many ways, including varying the amount of return on the public good both to the group and to the individual. For example, as cited by Goeree, Holt, and Laury (2002), research by Carter, Drainville, and Poulin (1992) investigated the possible differential returns—for example, they set an “internal return” for the public good (e.g., 5¢) and an external return for the public good (e.g., 2¢). In this way, the marginal value of the public good can also be contrasted to the net costs of contributing versus the benefit to others. This experimental strategy was also followed by Goeree, Holt, and Laury (2002) in an analysis of altruism in one-shot or one-time public goods settings.

### A Repeated Decisions versus One-Shot or One-Time Decisions

One of the most important theoretical differences in resource settings concerns whether or not the actors anticipate future interaction. In social psychological terms, anticipation of interaction carries with it many different factors. The interaction carries with it the ability of the other to respond, of course. Thus, if actor A contributes at time 1 while actor B does not contribute at time 1, actor A can retaliate against actor B at time 2; conversely, actors can also reinforce each other.

One-shot or one-time decisions do not carry the anticipation of future interaction and so also they do not admit the possibility to reward or punish other

actors. They are restrictive interactions, and cooperation should be, from all theoretical perspectives, difficult to maintain. Cooperation levels should be at the lowest in one-shot settings. Still, as noted previously, cooperation to the public or resource good is not zero.

In some experimental studies, one-shot decisions have been repeated. However, when the composition of the group is continuously shuffled so that the group members in decision 1 are completely different from the group members at decision 2, the claim is usually made that these are equivalent to one-shot decisions.

When groups are interacting over time, there are several theoretical conditions important to the establishment of the theoretical predictions and consequently the experimental design. If actors know the exact number of times that they will be interacting, they can engage in calculations associated with the endpoint. If actors know the endpoint, they realize that at that endpoint, there is minimal incentive to cooperate because other members will no longer have any sanctioning ability. This can lead to an “unraveling” of cooperation—or at least that has been the theory of “backward induction.” Although experimental research does not support the lack of cooperation or the extent of free-riding posited through a mechanism of backward induction (see a review of resource dilemmas by [Rapoport, 1997](#)), there are still problems with subjects knowing exactly how many trials will obtain.

Experimentalists have dealt with endpoint issues in several ways. Sometimes the number of trials is announced to the subjects, but the data for the endpoint are analyzed separately from the other data. There is some controversy about doing this because subjects do know the endpoint and this knowledge may create different behavior as the trials proceed. Another way this has been handled is for the experimenter to announce a range of trials that will occur and to then randomly select the number for a given experiment. This technique has the advantage of not having a specific endpoint but, practically speaking, gives subjects an idea of about how long they will be interacting.

At times, the amount of interaction is a very specific and important parameter of the study. According to what is usually called the “folk theorem,” contributing to a public good can be a “purely rational” individual strategy for an individual actor if the time horizon (or discount parameter) is sufficiently large. Basically, if an actor knows that the interaction is likely to occur for a long time, it is better to cooperate (as long as others are cooperating) because his or her share of the resource or public good could be greater. The statement “as long as others are cooperating” is an important theoretical proviso. The folk theorem is critical for theoretical development because it states that whereas cooperating at any one point in time is not individually rational, cooperating over longer periods of time can be justified without sacrificing any assumptions about rationality. However, although the folk theorem allows for the possibility of cooperation, it does not rule out very many possible outcomes, so additional assumptions are necessary for predicting likely outcomes.

If the discount parameter is either a scope condition of the theory or a manipulated portion of the study, it can be specifically implemented. Thus, for example, subjects can be “given” a projected number of trials. To eliminate the endpoint effects, a random component is added. In [Sell and Wilson \(1999\)](#), treatments involved several different discount parameters. It was explained to subjects that there was a likelihood or probability that each trial would be the last (this remained constant throughout the experiment). Subjects were told that this number represented the likelihood that the study would end and that another way to think about this was the likely number of trials. So, for example, if the likelihood was 0.95, that meant that the probability of this decision period being the last was 0.05 or 5 out of 100. Subjects were told, “Another way to think about this is as the average number of trials that you can expect to play. You figure this out by simply considering the fractions. If the probability was 0.95, then you could expect, on the average, to play 20 trials (5/100 or 1/20).” This probability was then concretely symbolized by chips in a basket. In the case of 0.95, for example, there might be 19 red chips and 1 white chip. After each decision, a chip was drawn to determine if the decision periods would continue.

In many resource goods experiments, the time dimension is controlled by the replenishment rate and, of course, the rate at which subjects might harvest the good. This replenishment rate can mirror the way in which biological resources such as fish might reproduce or, alternatively, die off from causes other than overfishing. New experimental paradigms to signify social–ecological systems have been developed to address both spatial and temporal resource dynamics ([Janssen, Holahan, Lee, & Ostrom, 2010](#)).

## V WHAT GROUP MEMBERS KNOW ABOUT EACH OTHER

One of the most important components for social dilemmas is what information actors have about each other. This can take the form of what characteristics of group members are revealed and what information about payoffs or motivations is available.

### A Symmetrical versus Asymmetrical Information

Information is symmetric if each actor knows that his or her information is the same as others. Asymmetric information is the instance in which some actors possess private information, or information that others do not have. This information can relate to the endowments that group members possess, the payoff structure, or characteristics of the actors. Research by [Sell and Wilson \(1991\)](#), for instance, demonstrated that actors contribute more to a public good when they have information about each individual group member’s contribution and when they know that others have this information as well.

Yamagishi and colleagues developed a manipulation to examine how a common identity might affect actors’ cooperation (or contributions) in a one-shot

prisoner's dilemma setting. This asymmetric information is used to determine the mechanism through which group identity might affect cooperation.

Group identity has been an important variable in the cooperation and trust literature. As discussed in [Yamagishi, Jin, and Miller \(1998\)](#) and [Yamagishi et al. \(2005\)](#), much research has been interpreted as supporting the general idea that actors give preference to those who share a group membership characteristic. A series of studies beginning with [Billig and Tajfel \(1973\)](#) and [Tajfel, Billig, Bundy, and Flamen \(1971\)](#) developed the concept of group differentiation based on a trivial label (such as preference between two artists—Klee or Kandinski—or whether an individual was likely to over- or underestimate the number of dots when exposed to slides containing a vast number of dots). Surprisingly, actors used this label to give preference to those with whom they shared the label and consequently gave them greater resources.

Yamagishi and colleagues conducted a series of experiments to examine the nature of this preference. To separate out the effects of favoritism built on the commonality versus the effects of expectations that other in-group members will acknowledge (and potentially reciprocate), the researchers developed a condition they called unilateral knowledge and compared its outcomes to those from mutual knowledge. Unilateral knowledge (a particular kind of asymmetrical knowledge) meant that a subject knew that the partner was in the same group as himself or herself; however, the subject also knew that the partner *did not* know. (See [Jin and Yamagishi, 1997](#).)

It was found that more cooperation was generated in settings in which there was mutual knowledge than unilateral knowledge. Such results support the idea that in-group favoritism is not just based on a shared membership; rather, it seems based on what Jin and Yamagishi called group heuristics, an expectation of generalized reciprocity from members of the same group. This general result was supported in the instance of a trivial group membership as well as the nationality (specifically, subjects knowing that their partner was either Australian or Japanese).

## B Simultaneous or Sequential Decisions

The implementation of particular decision-making rules can also affect the information that group members have about each other. One type of decision rule pertains to *when* group members make decisions relative to each other. In one-time decisions, if decisions are simultaneous, group members have no information about the others' decisions when they make their own decision. In sequential decisions, there is an order to the decision-making and, consequently, those who follow always have more information about others' decisions than the actors who preceded them.

[Yamagishi and Kiyonari \(2000\)](#) used the difference between simultaneous and sequential decision-making to further examine the impact of group identity. They manipulated group identification in terms of a trivial category—a given

group member was classified according to his or her preference for either Klee or Kandinski paintings. Subjects were then paired with a partner and told that they would make the first decision while their partner would make the second decision. In this particular manipulation of a dilemma, subjects received an endowment and had to decide how much of the endowment they would give to the partner. The money provided to the partner was doubled; then, subjects were told, the partner was faced with the same decision: how much of their endowment should they provide to the partner? (This is a very common experimental paradigm called the “trust game”.) Yamagishi and Kiyonari reasoned that if group identity was the only factor affecting decisions, there would be no difference between allocations made when decisions were simultaneous and when they were sequential. Their prediction, in line with the formulations discussed previously, was that cooperation would *not be* equivalent. That is, regardless of whether the group member was or was not a member of the same group, cooperation should be high in the sequential game. This, indeed, was the case and supported the idea that the hope of reciprocity prompted cooperation in the sequential game.

## C Common Knowledge

In economic investigations, in particular, the requirement of common knowledge is important for theoretical derivations and empirical application. This relates to some of the assumptions posited in game theory, the basis of much of the economic theory of social dilemmas. Basically, the common knowledge requirement is that everyone knows the parameters of the setting and everyone knows that everyone knows (Bagnoli & McKee, 1991). This terminology is a little different from that used about information symmetry and asymmetry. Whereas information usually refers to particular characteristics about actors or their past behavior, common knowledge usually refers to understandings about how others process information. Generally, this means that actors assume others have the information to act in a consistent (rational) manner. Thus, for example, a setting could be one in which there was asymmetric information but still be one in which there was common knowledge. To ensure that this requirement is met, it is usual to somehow illustrate to all subjects that everyone is reading and seeing the same information. (This information could include the idea that some subjects have information that others do not.) Although it is important that all subjects have their own sets of instructions and their own records of payoffs and earnings for interactions over time, for many studies it is also important for a researcher to read or demonstrate general study parameters. Along this line, it is a good idea to interview subjects over the relevant parameters of the studies and to answer any questions before beginning data collection. In this way, a researcher can assess that subjects believe that those with whom they will interact have the same information and understand the information in the same way as they do (see discussion in Holt & Davis, 1993).

## D Information about Characteristics of Group Members

As mentioned previously, characteristics of individuals can be used to examine group identity issues. Sometimes the group identity relates to contrived differences such as a particular personal style or ability that relates to over- or underestimating dots. Sometimes, the group identity is an actual characteristic such as nationality (Bahry, Kosolopov, Kozyreva, & Wilson, 2005; Yamagishi et al., 2005).

There has also been some investigation of other factors that might serve as indicators of how people might or might not cooperate. Some research, for example, has considered how knowledge of the sex category of those in the group affects cooperation (Sell, 1997; Sell, Griffith, & Wilson, 1993; Simpson, 2003). Some research has considered how racial or ethnic differences in the United States may be related to willingness to cooperate with a partner (Eckel & Wilson, 2004; Simpson, McGrimmon, & Irwin, 2007). Simpson, Willer, and Ridgeway (2012) explored the role of status characteristics in collective action necessary for solving social dilemmas. They found that higher status individuals often cooperate at higher levels than others in the group and can effectively influence those of lower status to also contribute. (These authors also cite the research of Sell (1997) and Kamru and Vesterlund (2010) as supporting their argument and interpretations.) To test their conjectures, these researchers created an experimental environment that featured an experimental paradigm in which participants were not presented with discrete decisions as in most studies but, rather, a real-time contribution decision such that any group member could decide to make a contribution before the allotted time expired. In this way, the researchers could manipulate the information supposedly relayed by other group members and also investigate when actual participants contributed.

Research also suggests that cues provided by group members, separate from individual characteristics, can affect the degree to which group members trust them or cooperate with them. Such cues include smiling (Scharlemann, Eckel, Kacelnik, & Wilson, 2001).

Methodologically important to these studies is the separation of the information alone from other content. Thus, information about sex category as carried through the gender-specific names of participants (e.g., “Michelle” vs. “Michael” as an indicator of sex category) is different from showing group members pictures of each other. Pictures carry other content as well, such as attractiveness, ethnicity, and general facial expression. It is not that one manipulation is correct while the other is not, but simply that the researcher must be aware of exactly what information is being conveyed.

## E Communication

Related to the nature of what information is available and how it is delivered is communication. A very consistent finding in the social dilemma literature

is that face-to-face communication among group members tends to increase cooperation. (See extended discussions of this consistency in [Sally \(1995\)](#) and [Ostrom \(1998\)](#).) The idea that communication alone helps solve dilemmas is contrary to much of noncooperative game theory because communication that is not binding (in the sense of contractual) is consequently considered “cheap talk.” However, despite the inability to enforce commitments, group communication still has a strong effect.

What is it about communication that increases cooperation? Researchers argue that communication creates commitments and often groups will try to encourage public commitment to a particular cooperative strategy ([Kerr, 1995](#); [Kerr & Kaufmann-Gilleland, 1994](#); [Orbell, van de Kragt, & Dawes, 1988](#); [Ostrom, Walker, & Gardner, 1992](#)) and that group discussion leads to in-group identity ([Brewer & Kramer, 1986](#)). Furthermore, there is evidence that after discussion, participants are better able to predict whether others would cooperate (see [Frank, Gilovich, & Regan, 1993](#)). It is also the case that, in repeated interactions, verbal monitoring and punishment can be invoked ([Ostrom et al., 1992](#)).

In Elinor Ostrom’s presidential address to the American Political Science Association (1998), she argued that communication was one of the most important methods for potential solutions to the social dilemma or collective action problem. In her analysis, she argues that face-to-face communication does many things to help provide solutions for social dilemmas. In particular, it creates an opportunity to develop “contingent agreements.” Essentially, these agreements state that one person or group is willing to contribute or cooperate at a particular level *if* others will also cooperate. Communication then enables each participant to estimate his or her degree of trust in those contingent agreements. Contingent agreements may be very simple (“Let’s all contribute equally”) or complex (“Let’s contribute according to how much our endowments are”). They may also deal with punishment for defection. After an initial decision, reputation and reciprocity become factors: reputation of group members is the result of how their communicated commitment relates to actual behavior, and reciprocity is the behavior that reflects matching prior behavior.

An important point about face-to-face communication made clear in [Ostrom’s \(1998\)](#) discussion is that face-to-face communication contains an incredible variety of factors. Because this is the case, it must also be the case that under certain conditions, face-to-face communication may *not* lead to solution of social dilemmas. Thus, for example, if a group member refuses to cooperate initially, this refusal can create problems for the group and recovery may be difficult. In Chapter 20 of this volume, Driskell, King, and Driskell make the point that gestures are important aspects of communication. “Simple communication” is hardly ever really simple.

But, indeed it is important. Some evidence indicates that just sending computerized signals as a form of communication does *not* increase cooperation; in fact, it really does seem to be cheap talk, not creating credible commitment ([Wilson & Sell, 1997](#)). So although the results of experimental research in this area seem

consistent, this also seems an important arena for trying to tease out what factors of communication are most important for determining trust under specific conditions. This is precisely where the experimental method has its strength.

## F Punishment and Triggers

An important aspect of social dilemmas involves the possibility of punishment. The folk theorem, mentioned in prior discussion, implies that there are an infinite number of ways that cooperation can be reinforced. One potential way is through the threat of a trigger. Trigger strategies are basically threats of noncontribution that would be “triggered” if others should demonstrate noncooperation. Evidence indicates that when the threat of a trigger that mandates defection forever given noncooperation of any group member (i.e., a “grim trigger”) exists, cooperation is higher than when it is not present (Sell & Wilson, 1999). The ability to punish select group members on an individual level also can be extremely influential. The experimental design features that capture theoretical properties of both reward and punishment are important; punishment or reward can vary in many different ways, such as in whether identities of both the givers and receivers of the punishment/reward are known and the costs of the punishments and rewards (Casari & Luini, 2009). As Fehr and Gächter (2000), Sefton, Shupp, and Walker (2007), Walker and Halloran (2004), Bigoni, Camera, and Casari (2013), and Casari and Luini (2009) investigate, under some conditions, punishment can be extremely effective. However, it is not always effective, and especially when everyone can punish without constraints, punishment can decrease cooperation (Casari & Luini, 2009).

The rules of the institution that enable punishment are important. In a study examining adoption of rules by participants, Ertan, Page, and Puttermann (2008) examined voting for rules that would allow punishment. They found that although participants did not often vote for allowing punishments for low contributors initially, there was a clear tendency to allow it over time. Those groups that did vote for and allow punishment achieved higher levels of cooperation than those that did not.

The issues of punishment, the identities of both those being punished and those doing the punishing, and the institutions and context surrounding the punishment are extremely active areas of research. Although the issue of punishment and reward has been extensively studied in sociological investigations of exchange (see Molm, Chapter 9, this volume), there is surprisingly little theoretical integration across economic, sociological, and social psychological models of punishment. This suggests an important new direction that would enable combining insights from different disciplines.

## VI NEW DIRECTIONS AND NEW STRATEGIES

There are some issues that serve to keep research on social dilemmas separated by discipline. In terms of experimental methodology, the use of deception

dramatically divides the disciplines. Economists are militantly opposed to the use of deception, whereas there are varying degrees of acceptance in psychology, political science, and sociology. The economists' argument made against deception relates to how subjects are changed by the deception and consequently what this means for the research results. This argument sometimes carries objections on ethical grounds, but for the most part, it is based on problems that deception may pose for ensuring that participants believe the experimenter. Thus, for example, if part of the research study related to different endowments and the subjects did not believe the information in the study, the results could not be interpreted.

Furthermore, because uncertainty is such an important concern in many economic investigations, subjects' suspicions about the truth of the information could dramatically affect the processes under investigation. Charles Holt (2007), an experimental economist, develops this argument using public goods language:

*Even if deception is successfully hidden during the experiment, subjects may learn the truth afterward by sharing their experiences with friends so the careful adherence to nondeceptive practices provides a public good for those who run experiments in the same lab at a later time. Ironically, the perverse incentive effects of deception in social psychology experiments may be aggravated by an ex post confession or debriefing, which is sometimes required by human subjects committees. (p. 12)*

Ortmann and Hertwig (1997, 1998, 2002), Hertwig and Ortmann (2001, 2002), and Devetag and Ortmann (Chapter 16, this volume) specifically discuss such issues and offer some creative suggestions for when and how to use deception in experimental designs. In particular, they argue against the routine use of deception or the use of deception for convenience. On the other hand, sociologists, psychologists, and political scientists often argue that deception, in the sense of providing false information to subjects, is necessary for certain types of research questions. In particular, questions that involve assessing the effects of false labels or false information, by definition, involve some deception.

Although both views have merits, the discussion of deception can be quite volatile. In practical terms, it is important to understand the views from the different research groups. Reviewers who are economists will not view experiments with deception favorably. It is especially important to discuss deception and design issues if any collaboration across disciplines is being considered. (See discussion in Cook and Yamagishi (2008), Hertwig and Ortmann (2008), and Sell (2008).)

There are also different theoretical approaches that can separate the disciplines studying social dilemmas. Although there is much overlap in the research issues investigated, differences in theoretical language and assumptions sometimes prevent productive accumulation. Game theory, so much a part of economics and political science, is not even taught in many graduate programs in sociology and psychology, for example.

On the other hand, in economics and political science, there has been a focused effort to delineate conditions predicted on the basis of rational choices theories, usually specified in terms of game theory. Thus, for example, as discussed previously, there was quite a bit of research oriented to developing a baseline to demonstrate the prediction of noncooperation or free-riding. When demonstrations of these predictions failed, there was a flurry of activity, some of which generated alternative game theoretic models (e.g., see [Benoit & Krishna, 1985](#)). Sociologists and psychologists, on the other hand, tended to emphasize the importance of norms, reciprocity, and group identity to examine the conditions under which participants were more or less likely to cooperate for a group. They were less affected than economists and political scientists by the lack of “rational action” evidence because their formulations tended to focus on empirical regularities rather than the purely mathematical foundations of rational choice.

There is a movement that seems to have the potential to bring together these two sets of disciplines. In particular, behavioral economics and behavioral political science are research movements that focus on a fuller integration of mathematical theory (particularly game theory) and insights from psychology and sociology. The goal here is to take the findings of research into account and to incorporate them into the traditional rational choice theories. Sociology and psychology were always “behavioral” in the sense that the theories of the disciplines had to withstand empirical test and incorporated understandings, and the advantage of teaming up with economics and political science comes from the strength of the mathematics in these fields. For example, expected utility theory provides an important foundation for many kinds of decisions. It enables predictions of future choices by observation of past choices between various alternatives to determine the relative value or utility of a choice. Furthermore, the mathematics implied by expected utility allow researchers to represent uncertainty as the likelihood of the event. Consequently, the probability of a particular choice for an actor is based on the likelihood or probability of the event multiplied by the value of the choice.

Another potential source of integration across these fields relates to examination of institutional rules. (For more discussion of this suggestion, see [Lovaglia, Mannix, Samuelson, Sell, and Wilson \(2005\)](#) and [Sell, Lovaglia, Mannix, Samuelson, and Wilson \(2004\)](#).) Institutional rules are general laws or principles that specify who may do what particular actions and when they are allowed to do them. For example, voting rules specify what constituency has the right to vote and what form that vote takes, such as majority rule. Institutional rules may be formal or informal, stable or unstable. (See [Crawford and Ostrom \(1995\)](#) for a discussion of the grammar of institutional rules.) Using this perspective and language enables comparison of groups that have vastly different purposes and composition. For example, institutional boundary rules delineate who is a member of the group and how a person might acquire that membership. In the case of social dilemmas, the defining property of nonexcludability—that

nobody may exclude anyone else from the group—for group members is an indication of the importance of boundary.

Very little experimental research has specifically considered the rules by which individuals or groups become members, but in-group literature suggests the importance of “common fate” in definitions of who is and is not a member. (For further discussion, see [Sell and Love \(2009\)](#).) Sometimes boundaries are permeable (members may leave and outsiders may become members), whereas under other conditions boundaries are impermeable. Conceptualizing the degree of permeability might be a way to examine the structure of strength of a social dilemma as well as its solution. For example, it may be that when boundaries are permeable, group identity is lower and consequently social dilemmas become more difficult to solve. Other institutional rules that have particular importance in the study of social dilemmas are information rules. (Some of these rules are discussed in the previous section on information.) *Information rules* describe how information is shared and what each actor can know. The information linkages determine the extent to which each member knows what the others have done or what they are planning to do. Information conditions can vary from complete information (everyone sees what everyone else has done to the point) to incomplete information or information characterized by uncertainty. The point here is that the language of institutional rules provides a method to connect across disciplines. It also provides a way to examine relations among nested groups or to highlight how institutions might reach from group to group.

The study of social dilemmas already reaches across discipline boundaries. In part, this is because there has been an acceptance of the experimental method. Future prospects for further interdisciplinary research seem even more promising if more common language develops to enable routine interaction and theoretical coordination.

## REFERENCES

- Andreoni, J. (1995). Warm-glow versus cold pickle: The effects of positive and negative framing on cooperation in experiments. *Quarterly Journal of Economics*, 110, 1–21.
- Bagnoli, M., & McKee, M. (1991). Voluntary contribution games: Efficient private provision of public goods. *Economic Inquiry*, 29, 351–366.
- Bahry, D., Kosolopov, M., Kozyreva, P., & Wilson, R. K. (2005). Ethnicity and trust: Evidence from Russia. *American Political Science Review*, 99(4), 521–532.
- Benoit, J. P., & Krishna, V. (1985). Finitely repeated games. *Econometrica*, 53, 905–922.
- Bigoni, M., Camera, G., & Casari, M. (2013). Strategies of cooperation and punishment among students and clerical workers. *Journal of Economic Behavior & Organization*, 94, 172–182.
- Billing, M., & Tajfel, H. (1973). Social categorization and similarity in intergroup behavior. *European Journal of Social Psychology*, 3, 27–52.
- Bohm, P. (1972). Estimating demand for public goods: An experiment. *European Economic Review*, 3, 111–130.
- Bowles, S., & Gintis, H. (2004). The evolution of strong reciprocity: Cooperation in heterogeneous populations. *Theoretical Population Biology*, 65, 17–28.

- Brewer, M., & Kramer, R. M. (1986). Choice behavior in social dilemmas: Effects of social identity, group size and decision framing. *Journal of Personality and Social Psychology, 50*, 543–549.
- Carter, J. R., Drainville, B. J., & Poulin, R. P. (1992). *A test for rational altruism in a public goods experiment*. Worcester, MA: Working paper, College of the Holy Cross.
- Casari, M., & Luini, L. (2009). Cooperation under alternative punishment institutions: An experiment. *Journal of Economic Behavior & Organization, 71*, 273–282.
- Cook, K., & Yamagishi, T. (2008). A defense of deception on scientific grounds. *Social Psychology Quarterly, 71*, 215–221.
- Crawford, S. E. S., & Ostrom, E. (1995). A grammar of institutions. *American Political Science Review, 89*, 582–600.
- Dawes, R. M. (1980). Social dilemmas. *Annual Review of Psychology, 31*, 169–193.
- Eckel, C., & Wilson, R. K. (2004). Is trust a risky decision? *Journal of Economic Behavior and Organization, 55*, 447–466.
- Ertan, A., Page, T., & Puterman, L. (2008). Who to punish? Individual decisions and majority rule in mitigating the free rider problem. *European Economic Review, 53*, 495–511.
- Fehr, E., & Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review, 90*, 980–994.
- Fisher, J., Isaac, R. M., Schatzberg, J., & Walker, J. (1995). Heterogeneous demand for public goods: Effects on the voluntary contributions mechanism. *Public Choice, 85*, 249–266.
- Frank, R. H., Gilovich, T., & Regan, D. T. (1993). The evolution of one-shot cooperation: An experiment. *Ethology and Sociobiology, 14*, 247–256.
- Goeree, J. K., Holt, C. A., & Laury, S. K. (2002). Private costs and public benefits: Unraveling the effects of altruism and noisy behavior. *Journal of Public Economics, 83*, 257–278.
- Hardin, G. (1968). The tragedy of the commons. *Science, 162*, 1243–1248.
- Hertwig, R., & Ortmann, A. (2001). Experimental practices in economics: A methodological challenge for psychologists? *Behavioral and Brain Sciences, 24*, 383–451.
- Hertwig, R., & Ortmann, A. (2002). Economists' and psychologists' experimental practices: How they differ, why they differ and how they could converge. In I. Brocas, & J. D. Carillo (Eds.), *The psychology of economic decisions* (pp. 253–272). New York: Oxford University Press.
- Hertwig, R., & Ortmann, A. (2008). Deceptions in social psychological experiments: Two misconceptions and a research agenda. *Social Psychology Quarterly, 71*, 222–227.
- Holt, C. A. (2007). *Markets, games, & strategic behavior*. Boston: Pearson.
- Holt, C. A., & Davis, D. (1993). *Experimental economics*. Princeton, NJ: Princeton University Press.
- Isaac, R. M., & Walker, J. (1988). Group size effects in public goods provision: The voluntary contributions mechanism. *Quarterly Journal of Economics, 103*, 179–199.
- Isaac, R. M., Walker, J., & Thomas, S. (1984). Divergent evidence on free riding: An experimental examination of possible explanations. *Public Choice, 43*, 113–149.
- Janssen, M. A., Holahan, R., Lee, A., & Ostrom, E. (2010). Experiments for the study of social-ecological systems. *Science, 328*, 613–617.
- Jin, N., & Yamagishi, T. (1997). Group heuristics in social dilemma. *Japanese Journal of Social Psychology, 12*(3), 190–198.
- Kamru, C. S., & Vesterlund, L. (2010). The effect of status on charitable giving. *Journal of Public Economic Theory, 12*, 709–735.
- Kerr, N. (1995). Norms in social dilemmas. In D. Schroeder (Ed.), *Social Dilemmas: Perspectives on Individuals and Groups* (pp. 31–47). Westport, CT: Praeger.
- Kerr, N., & Kaufman-Gilleland, C. (1994). Communication, commitment, and cooperation in social dilemmas. *Journal of Personality and Social Psychology, 66*(3), 513–529.

- Kollock, P. (1998). Social dilemmas: The anatomy of cooperation. *Annual Review of Sociology*, 24, 183–214.
- Kollock, P., & Smith, M. (1996). Managing the virtual commons: Cooperation and conflict in computer communities. In S. Herring (Ed.), *Computer-mediated communication: Linguistic, social, and cross-cultural perspectives* (pp. 109–128). Amsterdam: Benjamins.
- Ledyard, J. (1995). Public goods: A survey of experimental research. In J. H. Kagel, & A. E. Roth (Eds.), *Handbook of experimental economics* (pp. 111–194). Princeton, NJ: Princeton University Press.
- Lovaglia, M., Mannix, E. A., Samuelson, C. D., Sell, J., & Wilson, R. K. (2005). Conflict, power and status in groups. In M. S. Poole, & A. B. Hollingshead (Eds.), *Theories of small groups: Interdisciplinary perspectives* (pp. 139–184). Thousand Oaks, CA: Sage.
- Marwell, G., & Ames, R. E. (1979). Experiments on the provision of public goods: I. Resources, interest, group size and the free-rider problems. *American Journal of Sociology*, 84, 1335–1360.
- Marwell, G., & Ames, R. E. (1980). Experiments on the provision of public goods: II. Provision points, stakes, experience and the free rider problem. *American Journal of Sociology*, 85, 926–937.
- Marwell, G., & Ames, R. E. (1981). Economists free ride, does anyone else? Experiments on the provision of public goods IV. *Journal of Public Economics*, 15, 295–310.
- McCabe, K. A., & Smith, V. L. (2005). Goodwill accounting and the process of exchange. In G. Gigerenzer, & R. Selten (Eds.), *Bounded rationality* (pp. 319–340). Boston: MIT Press.
- McCusker, C., & Carnevale, P. J. (1995). Framing in resource dilemmas: Loss aversion and the moderating effects of sanctions. *Organizational Behavior and Human Decision Processes*, 61, 190–201.
- Messick, D. M., Allison, S. T., & Samuelson, C. D. (1988). Framing and communication effects on groups members' responses to environmental and social uncertainty. In S. Maital (Ed.), *Applied behavioral economics*, Vol. 2. (pp. 677–700). New York: New York University Press.
- Nowak, M. A. (2006). Five rules for the evolution of cooperation. *Science*, 314, 1560–1563.
- Olson, M., Jr. (1965). *The logic of collective action: Public goods and the theory of groups*. Cambridge, MA: Harvard University Press.
- Orbell, J. M., van de Kragt, A. J. C., & Dawes, R. M. (1988). Explaining discussion-induced cooperation. *Journal of Personality and Social Psychology*, 54, 811–819.
- Ortmann, A., & Hertwig, R. (1997). Is deception acceptable? *American Psychologist*, 52, 746–747.
- Ortmann, A., & Hertwig, R. (1998). The question remains: Is deception acceptable? *American Psychologist*, 53, 806–807.
- Ortmann, A., & Hertwig, R. (2002). The empirical costs of deception: Evidence from psychology. *Experimental Economics*, 5, 111–131.
- Ostrom, E. (1998). Rational choice theory of collective action. *American Political Science Review*, 92, 1–22.
- Ostrom, E., Walker, J., & Gardner, R. (1992). Covenants with and without a sword: Self-governance is possible. *American Political Science Review*, 86, 404–417.
- Rapoport, A. (1997). Order of play in strategically equivalent games in extensive form. *International Journal of Game Theory*, 26, 113–136.
- Rutte, C. G., Wilke, H. A. M., & Messick, D. M. (1987). The effects of framing social dilemmas as give-some or take-some games. *British Journal of Social Psychology*, 26, 103–108.
- Sally, D. (1995). Convention and cooperation in social dilemmas: A meta-analysis of experiments from 1958 to 1992. *Rationality and Society*, 7, 58–92.
- Samuelson, P. A. (1954). The pure theory of public expenditure. *The Review of Economics and Statistics*, 36, 387–389.
- Samuelson, P. A. (1955). Diagrammatic exposition of a theory of public expenditure. *The Review of Economics and Statistics*, 37, 350–356.

- Samuelson, P. A. (1958). Aspects of public expenditure theories. *The Review of Economics and Statistics*, 40, 332–338.
- Scharlemann, J. P., Eckel, C. C., Kacelnik, A., & Wilson, R. K. (2001). The value of a smile: Game theory with a human face. *Journal of Economic Psychology*, 22, 617–640.
- Schwartz-Shea, P., & Simmons, R. T. (1995). Social dilemmas and perceptions: Experiments on framing and inconsequentiality. In D. A. Schroeder (Ed.), *Social dilemmas: Perspectives on individuals and groups* (pp. 87–103). Westport, CT: Praeger.
- Sefton, M., Shupp, R., & Walker, J. M. (2007). The effect of rewards and sanctions in provision of public goods. *Economic Inquiry*, 45(4), 671–690.
- Sell, J. (1988). Types of public goods and free-riding. *Advances in Group Processes*, 5, 119–140.
- Sell, J. (1997). Gender, strategies, and contributions to public goods. *Social Psychology Quarterly*, 60, 252–265.
- Sell, J. (2008). Introduction to deception debate. *Social Psychology Quarterly*, 71, 213–214.
- Sell, J., Chen, Z. Y., Hunter-Holmes, P., & Johansson, A. C. (2002). A cross-cultural comparison of public good and resource good settings. *Social Psychology Quarterly*, 65, 285–297.
- Sell, J., Griffith, W. I., & Wilson, R. K. (1993). Are women more cooperative than men in social dilemmas? *Social Psychology Quarterly*, 56, 211–222.
- Sell, J., Lovaglia, J. J., Mannix, E. A., Samuelson, C. A., & Wilson, R. K. (2004). Investigating conflict, power and status within and among groups. *Small Group Research*, 35, 44–72.
- Sell, J., & Love, T. P. (2009). Common fate, crisis and cooperation in social dilemmas. *Advances in Group Processes*, 26, 53–79.
- Sell, J., & Son, Y. (1997). Comparing public goods with common pool resources: Three experiments. *Social Psychology Quarterly*, 60(2), 118–137.
- Sell, J., & Wilson, R. K. (1991). Levels of information and public goods. *Social Forces*, 70, 107–124.
- Sell, J., & Wilson, R. K. (1999). The maintenance of cooperation: Expectations of future interaction and the trigger of group punishment. *Social Forces*, 77, 1551–1570.
- Simpson, B. (2003). Sex, fear, and greed: A social dilemma analysis of gender and cooperation. *Social Forces*, 82, 35–52.
- Simpson, B., McGrimmon, T., & Irwin, K. (2007). Are blacks really less trusting than whites? Revisiting the race and trust question. *Social Forces*, 86, 525–552.
- Simpson, B., Willer, R., & Ridgeway, C. (2012). Status hierarchies and organization of collective action. *Sociological Theory*, 30, 149–166.
- Son, Y., & Sell, J. (1995). Are the dilemmas posed by public goods and common pool resources the same? *Advances in Human Ecology*, 4, 69–88.
- Tajfel, H., Billig, M. G., Bundy, R. P., & Famen, T. C. (1971). Social categorization and intergroup behavior. *European Journal of Social Psychology*, 11, 439–443.
- Walker, J. M., & Halloran, M. A. (2004). Rewards and sanctions and the provision of public goods in one-shot settings. *Experimental Economics*, 7(3), 235–247.
- Wilson, R. K., & Sell, J. (1997). Cheap talk and reputation in repeated public goods settings. *Journal of Conflict Resolution*, 41, 695–717.
- Yamagishi, T., Jin, N., & Miller, A. S. (1998). In-group bias and culture of collectivism. *Asian Journal of Social Psychology*, 1, 315–328.
- Yamagishi, T., & Kiyonari, T. (2000). The group as the container of generalized reciprocity. *Social Psychology Quarterly*, 63, 116–132.
- Yamagishi, T., Makimura, Y., Foddy, M., Matsuda, M., Kiyonari, T., & Platow, M. (2005). Comparisons of Australians and Japanese on group-based cooperation. *Asian Journal of Social Psychology*, 8, 173–190.

## Chapter 11

# Hypotheses, Operationalizations, and Manipulation Checks

Martha Foschi

*University of British Columbia, Vancouver, British Columbia, Canada*

## I INTRODUCTION

The formulation and empirical test of hypotheses are two fundamental activities in scientific research. In this chapter, I examine some aspects of these activities within the context of experimental work in the social sciences and, in particular, laboratory experiments in group interaction. I focus on how the empirical testing of theoretical hypotheses requires both that their abstract terms be translated into concrete operations and that the effectiveness of these operations be verified.

Because the term “hypothesis” has many meanings, it is useful to begin by presenting a definition. By a hypothesis, I mean a statement, intended to be tested empirically, that proposes how a set of variables (or factors or features) are related. An example would be “the more knowledgeable a person appears to be, the more he or she will be liked by others.” The hypothesis states a positive or direct relationship between an independent variable, “perceived level of knowledge,” and a dependent (or outcome) variable, “level of liking.” It is, however, a very simplistic hypothesis because it ignores other factors that most probably also affect that relationship.

The hypothesis could become both more complex and more likely to be true if one were to reformulate it as follows: “In groups where members have no history of interaction with each other and where they share a strong motivation to solve a common task, the more knowledgeable a group member appears to be about that task, the more he or she will be liked by the other group members.” The hypothesis has now acquired four *scope conditions* (some stated more explicitly than others). These conditions indicate circumstances that define the applicability of the hypothesis and are as follows: a group with a common goal; participants with no history of interaction with each other; a strong motivation to achieve that goal; and communicated knowledge that is perceived to be related to the task (rather than knowledge about any matter). Other scope conditions could also be added.

Moreover, the hypothesis could be made more complex by including other types of variables. Thus, an additional *independent* variable could specify that the relationship will be affected by whether the person communicates his or her task-related knowledge either directly or indirectly. The more direct the expression of that knowledge, the more pronounced its effect on liking will be. An added *dependent* variable could stipulate that not only liking but also degree of influence exerted by that person will be simultaneously affected. Finally, an *intervening* variable such as “degree of respect toward that group member’s opinions” could mediate the relationship between the independent and the dependent variables.<sup>1</sup> A hypothesis can, and often should, have a number of each of these components. (On the different types of variables that make up a hypothesis, see [Baron & Kenny, 1986](#); [Foschi, 1997](#).)

The variables in a hypothesis may range from notions that summarize everyday experiences (e.g., “book” and “food”) to constructs that are part of a theory (e.g., “cognitive dissonance” or “social attribution”). The hypothesis itself, in turn, may be proposed as a stand-alone generalization or as a derivation from a scientific theory (here referred to as “theory”).

A theory is a system of ideas consisting of: (1) defined and undefined concepts that incorporate selected variables; (2) logically interconnected statements (assumptions and derived hypotheses) that link those concepts to each other; and (3) rules for the derivation of statements (see, e.g., [Cohen, 1989](#), Chapter 10; [Jasso, 1988](#); [Wagner, 1984](#), Chapter 2). Such a theory constitutes both an explanation of and a prediction about the occurrence of a class (or classes) of phenomena; a successful theory fulfills both of these functions—explanation and prediction—while exhibiting firm empirical support for its hypotheses.

A related, useful distinction is to classify research as either “exploratory” or “hypothesis testing.” In the former case, a set of variables is identified, but no relationship between them is proposed; in the latter, such a relationship is advanced. Hypothesis-testing studies that are embedded in a theory have a much larger long-term payoff than exploratory work that is not part of a theory. Whereas such exploratory work may or may not produce interesting findings, theoretical hypothesis-testing studies always yield useful outcomes: in the latter studies, there is a more clearly defined context within which the results can be interpreted—that is, they either give full, partial, or no support to the hypothesis. If full support is not obtained, the findings can point to flaws in either the hypothesis or the design. Moreover, assuming an appropriate design, results of any level of support are informative not only about the hypothesis but also about

---

1. This hypothesis is meant only as an example of how these variables *could* be related. Thus, rather than considering “communication (either direct or indirect) of knowledge about the group task” as an additional independent variable, “a direct communication of that knowledge” could be treated as a scope condition. In turn, a strong motivation to solve the common task could be a value of an independent variable spanning from “very strong” to “very weak” in that motivation. Several other models could also be proposed.

the theory. Results from atheoretical exploratory studies, on the other hand, stand by themselves. The absence of a theory means that there are no guidelines to sort out their various (often many) possible interpretations. For a useful, related distinction, see [Zelditch \(2013, pp. 5–6\)](#) on “theory-testing experiments” and “effect-oriented experiments.”

Next, I propose a set of theoretical hypotheses and construct an experimental test for them. Through this example, I put forth in concrete terms my thoughts about hypotheses and design issues. The points I make, however, are intended to apply beyond the specific details I present in this chapter.

## II HYPOTHESES

Suppose that a person had received results from a test of logical skills, and that these results were very poor. Would the person quickly conclude that he or she does not possess those skills? The answer would depend on several factors, such as the extent to which the results are accepted to be objective evaluations and how tired the person was at the time of the exam. Here, I focus on the perceived diagnosticity of the test and formulate my example within the context of expectation states theory, a system of ideas that meets all the requirements for a scientific theory listed previously.

The central interest of expectation states research is on how members of task groups assign competence<sup>2</sup> to each other ([Berger, Fisek, Norman, & Zelditch, 1977](#); for reviews, see [Berger & Webster, 2006](#); [Bianchi, 2010](#); [Correll & Ridgeway, 2003](#)). Two of the core concepts are “status characteristics” and “performance expectations.” The former is defined as any socially valued attribute seen by individuals as implying task competence. Such characteristics consist of at least two levels or “states” (e.g., either high or low mechanical ability), one of which is viewed as having more worth than the other. “Performance expectations” are beliefs about the likely quality of group members’ future performances on the task at hand, and they reflect levels of assigned competence. These beliefs need not be conscious or have an objective validity.

Status characteristics are classified as varying from specific to diffuse, depending on the perceived range of their applicability. A specific characteristic is associated with performance expectations in a limited domain; a diffuse characteristic carries expectations about performance on a wide, indeterminate set of tasks. For instance, in many societies, gender, ethnicity, nationality, formal education, and socioeconomic class constitute diffuse status characteristics for large numbers of individuals and in a variety of settings. Expectation states are said to develop for “self” (the focal person) relative to each of the other members of the group; all propositions are formulated from self’s point of view.

---

2. I treat “competence,” “skill,” and “ability” as synonyms. I also use “participants,” “group members,” “performers,” “respondents,” “self and partner” as equivalent notions.

The theory specifies how, and under what conditions, expectations are formed. For example, these may be based on the group members' status characteristics, on the evaluations they receive on their task performances by a "source" outside the group, or on both. Two central scope conditions are that members are "task oriented" (i.e., they value the ability required to do the task well and are motivated to achieve that result) and "collectively oriented" (i.e., they are prepared to both consider and use the other group members' opinions for the solution of the task).

Expectation states theory is, in fact, a set of interrelated theories or "branches." In this chapter, I focus on one of the main branches—namely, evaluations-and-expectations theory. This branch concerns the processes through which evaluations of units of performance generalize to become stable beliefs about levels of competence and result in performance expectations. Within this branch, research has considered properties of the performers and of the evaluations and has tested for the effects of these variables on expectation formation (see, e.g., [Webster & Sobieszek, 1974](#)). My interest here is in one feature of the evaluations: the extent to which they are seen as representing a diagnostic test of the ability in question. By "diagnosticity," I mean the degree to which a sample of a person's performances on a given task is considered to be a valid indication of his or her overall ability for that task. My general prediction is that, for good as well as poor levels of performance, a test believed to be highly diagnostic results in more conclusive (i.e., certain) inferences of either ability or lack of ability than does a less diagnostic test. In turn, I relate these inferences to levels of influence rejection, as I describe later.

This research topic is closely related to the work presented in [Foschi, Warriner, and Hart \(1985\)](#). In that study, we investigated the role that standards for both competence and lack of competence play in the formation of expectations. Here, I formulate my hypotheses for the same setting as in that study (which is also the setting investigated in a large proportion of expectation states research, as discussed in Berger, Chapter 12, this volume). That situation involves: (1) two persons, self and a partner ("other"), who perform a task, first individually and then as a team; and (2) a source of performance evaluations. The following scope conditions apply:

- (a) The task consists of a series of trials, each having the same level of difficulty.  
Self has no prior expectations about the ability required for this task and believes this ability to be both valuable and specific.
- (b) Self is motivated to perform the task well (i.e., self is "task oriented").
- (c) Performance evaluations originate in a source—namely, a person or procedure considered by self to be more capable of evaluating performances than is he or she. The source is the only basis of evaluations available to self to judge his or her task ability relative to that of the partner (and to form corresponding expectations).

(d) During the team phase of the task, self is prepared to both take into account and use the partner's ideas for the task solution (i.e., self is "collectively oriented").<sup>3</sup>

I propose the following hypotheses:

- If self has received *better* evaluations than the partner in the individual phase, then he or she will form *higher* expectations for self than for the partner and, in the team phase, will reject *more* influence from that person than if self had received *no* performance evaluations.
- If self has received *worse* evaluations than the partner in the individual phase, then he or she will form *lower* expectations for self than for the partner and, in the team phase, will reject *less* influence from that person than if self had received *no* performance evaluations.
- *Higher expectations* for self than for other will be *more conclusive* and result in *more* rejection of influence when the task has been perceived to be high rather than low in diagnosticity.
- *Lower expectations* for self than for other will be *more conclusive* and result in *less* rejection of influence when the task has been perceived to be high rather than low in diagnosticity.

In addition to the scope conditions, the hypotheses contain four variables: two independent (task diagnosticity and level of performance evaluations); one intervening (performance expectations); and one dependent (influence rejection). To my knowledge, these hypotheses have not yet been investigated within the expectation states tradition. It is also worth noticing that they implicitly contain what I call *irrelevant variables of theoretical interest*. These are factors that could reasonably be proposed to have a theoretical impact on the dependent variable(s) but that nevertheless have not been included (i.e., are treated as irrelevant) (Foschi, 1980, p. 93). For example, relative to self, the source could be of various levels of perceived superiority in capacity to evaluate performances (e.g., "marginally superior" and "clearly superior"). As an initial approach to the topic, I have assumed that those levels do not affect the results. Similarly, I am treating as theoretically irrelevant whether or not, across the trials, the distribution of evaluations favoring one performer over the other follows a pattern.

Every researcher has to make decisions about identifying irrelevant features. In some cases, these factors *are known* to be irrelevant to the process under

---

3. Note that some scope conditions are required at the beginning of the interaction as well as throughout (e.g., that the participant be task oriented), whereas others are required only at the beginning (e.g., that, at first, the person have no other bases for self-other expectations). Sometimes, the latter conditions are called "initial." I, however, prefer Cohen's terminology (1989, p. 80) and reserve the expression "initial conditions" for singular statements describing a particular system at a particular time and place. An initial condition is thus a statement that a scope condition has been operationalized as intended (see also Foschi (1997) and Section V in this chapter).

study (as supported by previous research). Other times, there is insufficient knowledge about a topic; in that case, one takes educated risks as to what factors can be left out. Finally, in still other cases, a variable can be assumed to be irrelevant *for the time being* (e.g., on the basis of what would be feasible to test). Regardless of the reasons, irrelevant features are seldom explicitly listed; this is done only if one wants to call special attention to the factor. Rather, it is generally understood that a theory (or a set of hypotheses within a theory) need not include all factors of possible interest; it is sufficient that only the key ones be listed.<sup>4</sup> If, however, a factor is deemed to be irrelevant though of theoretical interest, then it is included in the hypotheses through a clause specifying that they hold “regardless of the level of this variable.” The same as scope conditions, such theoretically irrelevant variables help define the class (or classes) of phenomena to which a hypothesis applies and are thus an instrument in generalizing it. I discuss irrelevant variables in more detail later in this chapter.

Let us now design an experimental test of the hypotheses.

### III THE EXPERIMENT

For research cumulativeness, work on expectation states has used a standardized experimental setting for hypothesis testing. The setting was created by Joseph Berger ([Berger et al., 1977](#), Chapter 5; [Camilleri & Berger, 1967](#); for a review of its history, see Chapter 12, this volume), and it has been used in several dozens of studies. In my test, I propose to use the computerized version that I have introduced in [Foschi \(1996\)](#). This version enables the researcher to re-create the same theoretical variables as does the original setting, and it has been an effective instrument for the study of expectation formation (for examples of other work using this version, see [Foschi, Enns, & Lapointe, 2001](#); [Lovaglia & Houser, 1996](#)). The following summarizes the main components of the procedures I propose for my test.

The experiment will be conducted over a series of sessions; two previously unacquainted persons (either two men or two women) will take part in each. They will be either first- or second-year undergraduates at my school, the University of British Columbia, and their ages will range between 18 and 21 years. They will be volunteers for a “study in visual perception” and will be recruited from large undergraduate classes. Each person will be paid the standard rate at this university (currently approximately \$16) for his or her participation, which will last approximately 1 hour and 20 minutes. The sessions will take place in a specially equipped research facility at the university. The two persons will be seated individually at adjacent computer stations said to be linked to each other. The stations will be separated by a fixed partition, and the participants will be precluded from both

---

4. For simplicity, here I list only a few factors in my hypotheses. Thus, I am not relating the hypotheses to other expectation states work that identifies variables (e.g., the strictness of performance standards and the number of evaluations) that elaborate and expand on how evaluations are interpreted.

seeing and talking with each other before or during the session; the only information they will be given about self and partner is that they are of the same sex category and year at the university and of similar age.

In order to keep sex category of experimenter constant, this person will be a female research assistant who will introduce herself as a graduate student in sociology (a statement that will be supported by her professional and confident demeanor); she will also be slightly older than the participants. She will read the instructions from a position that will enable her to make eye contact with the two persons; a summary of her statements will also appear on their computer monitors as the instructions are being read. The experimenter will remain in the room throughout the session, and her presence will be visible to the two participants at all times.

She will inform each pair that they will be asked to solve “contrast sensitivity” problems. This task involves viewing a series of rectangles made up of smaller ones (either red or white) that form an overall abstract pattern. Participants have to decide which color is predominant in each case. Although the proportion of the two colors is almost the same in all patterns, their configurations (different for each trial) create the impression that discriminating between the colors is a possible but difficult task. The high ambiguity in the color proportions, and the limited time (a few seconds) that the performers have for each decision, make it impossible for them to complete their task with certainty (e.g., by counting the colored rectangles). The experimenter will, however, state that “reliable research has established contrast sensitivity to be a newly discovered, important, and mainly innate ability.” She will also mention that that research has so far determined it to be relatively specific—that is, not related to attributes such as sex category, intelligence, or artistic skills.

An overview of the design and predictions is shown in [Table 11.1](#). Participants will *first* work individually on a series of 20 patterns and then will receive the “scores” obtained by self and partner. These will be communicated through each person’s monitor as well as printouts. The scores will be either 17 for self and 3 for the partner or the opposite combination. The two will also be informed of the trial-by-trial results for self and partner; these results will show one person consistently (although not on all the trials) outscoring the other. In addition, one group of control participants will perform the task but will receive no scores.<sup>5</sup> Pairs will be assigned at random to one of these three conditions. In turn, those receiving scores will be assigned, also at random, to hear a description of the task as either high or low in ability to diagnose contrast sensitivity reliably. The study thus involves five conditions or groups: one control (or baseline) and four experimental (or treatment).

---

5. Note that other control groups are also possible (and indeed desirable). For example, in one such group, the two performers could receive equally average (or good or poor) evaluations; in another, the first phase could be omitted altogether. In my view, the control group that I have included will provide the most basic information needed in this case.

**TABLE 11.1** Overview of the Experiment

Condition	Phase I			Rejection of influence measured*	
	Test score (out of 20) received by		Diagnosticity of test		
	Self	Other			
(1) Conclusive higher expectations for self than for other	17	3	High	a	
(2) Inconclusive higher expectations for self than for other	17	3	Low	b	
(3) Uninformed expectations	No information		No information	c	
(4) Inconclusive lower expectations for self than for other	3	17	Low	d	
(5) Conclusive lower expectations for self than for other	3	17	High	e	

\*Hypotheses: a > b > c > d > e.

Next, in all conditions, each pair will be instructed to work as a team on 25 patterns of a similar task said to require the same ability. This phase will be presented to them as an additional test of contrast sensitivity. On each of these trials, respondents will first make an initial choice between the two colors, receive the partner's "choice" via the monitor, and then make a final decision (either agreeing or disagreeing with him or her).<sup>6</sup> At the end, each person will individually complete a written questionnaire and will then be interviewed and debriefed. The questionnaire and interview will provide key manipulation checks.

6. The prearranged messages sent to the pair through their respective monitors enable the experimenter to assign two persons to the same condition in each session. For example, both can be informed simultaneously that self has received a higher score than the partner and that the task is high on diagnosticity. Moreover, in all five conditions, each person is informed *on the same prearranged trials* that the partner either agrees or disagrees with self's choices.

## IV OPERATIONALIZATIONS AND MANIPULATIONS

An operationalization is the translation of a theoretical variable into procedures designed to give information about its levels. I treat “operationalization,” “operational definition,” “measure,” “indicator,” and “observable” as close notions. Operationalizations tie theoretical ideas to evidence. For example, in the experiment proposed here, the case of “better evaluations for self than for the partner” becomes a score of 17 for one person and 3 for the other. My interest in this chapter is in the logic of the relationship between theoretical variables and their operationalizations; issues such as how to construct the measures, how to pretest them to ensure their validity and reliability, and what sample size to select are outside the scope of my discussion.

The notion of “manipulating a variable” overlaps in some respects with that of an operationalization. In an experiment, the researcher has a high degree of control over the variables that are either predicted or suspected to affect the participants’ responses. Ideally, all except the outcome variables (namely, the intervening and the dependent variables) are controlled through manipulation. Although the term “manipulation” is often used to refer only to the creation of levels of the independent variable(s) at the operational level, I interpret the term more broadly, as follows. *Direct* manipulation involves both establishing specific values of different types of variables (e.g., the belief that contrast sensitivity is a valuable ability, or that the test has either high or low diagnosticity) and keeping factors such as the participants’ age either constant or within a narrow range.

With respect to other factors that are not under study—and often not even identified as having potential relevance to the research (e.g., personality attributes of the participants or their mood during the experimental session)—control is exerted *indirectly* through the use of random assignment either to experimental or to baseline conditions. Random assignment tends to distribute these other factors evenly across those groups and thus ensures that they are not a variable affecting the responses.<sup>7</sup> It is useful to think of an experiment as a stage (or framework or set of constraints) within which several factors are controlled and a few (the intervening and dependent variables) are allowed to vary freely.

At this point, it is also helpful to make the following distinction. The independent variables of a hypothesis may be factors that are established by the researcher, such as the levels of performance evaluations and test diagnosticity described previously, and for which random assignment can be implemented. These are *truly experimental variables*. On the other hand, sometimes naturally occurring differences in characteristics of either the participants or the

---

7. Even if the distribution of such a factor is the same across conditions, the mean and/or the shape of the distribution could still affect the results. This issue can only be addressed through further work in which the factor is singled out and purposely varied across studies.

experimental situation (or both) are also of interest to the researcher, such as when he or she considers participants' sex category or location of the study as possible independent variables. Since random assignment cannot take place in these cases, the variables are known as *quasi-experimental* (or "organismic" because they are an inherent part of the entity under study). Many statistical tests of significance assume random assignment to conditions by the experimenter; it is only by accepting a relaxation of this requirement that those tests are used with quasi-experimental factors. For a valuable discussion of this subject, see [Magnusson and Marecek \(2012, p. 174\)](#).

## V ON PARTICULAR FEATURES

Setting the stage in order to test a theoretical hypothesis experimentally involves many design decisions. These reflect a combination of: (1) decisions about how to operationalize the theoretical variables of interest; (2) requirements dictated by good experimental design (e.g., random assignment to conditions); and (3) commonsense, practical matters (e.g., the length of the experimental session). All of these decisions contribute to narrowing down the theoretical ideas to specific situations. While all theoretical hypotheses are abstract, all evidence is concrete—that is, bound by particular circumstances, including those of time and place. Thus, such hypotheses always contain a larger number of cases than does the evidence about them. It is also important to consider that, even in a carefully constructed study, design decisions other than the ones taken could always have been made. Both in theoretical formulation and in test design, research is a continuously changing and self-improving activity.

In what follows, I examine the design decisions in my proposed experiment in relation to three types of particular features: (1) those that operationalize theoretical variables; (2) those that are truly irrelevant; and (3) those that represent experimental limitations. Note that the intervening variable of performance expectations is viewed in this theory as an unobservable construct and therefore is not measured directly in the standardized experimental setting ([Berger et al., 1977, p. 19](#)).<sup>8</sup>

### A Operationalization of Theoretical Variables

#### 1 Scope Conditions

Scope condition (a) is implemented through the features of the contrast sensitivity task. All patterns making up this task have the same high level of ambiguity, which performers are expected to perceive as high difficulty. The number of patterns in

---

8. Expectation states theory makes predictions about participants' behaviors, not about their self-reports. Notice, however, that some authors (see, e.g., [Foddy & Smithson, 1999](#); [Foschi, 1996](#)) have included such reports of self-other expectations in their designs as either auxiliary measures or manipulation checks. See also my comments, later in this chapter, on behaviors and self-reports.

each phase of the session reflects the following considerations: a much smaller number would probably be seen as inadequate for a test; a much larger number would likely make participants lose interest. In addition, the ambiguity of the task and its description as involving a specific and newly discovered ability make it very unlikely that those taking part would hold prior self-other expectations. The ability is also described as valuable. This statement is supported by the fact that the session takes place at a university facility, the reference by the experimenter to “reliable research on contrast sensitivity,” the perception that the current project is part of such research, and the computerized nature of the task. It is predicted that the latter feature will serve to associate the project with up-to-date technological developments and thus add interest and prestige to it. (For a further discussion of these points, see Foschi, 1997, pp. 541–543.)

The operationalization of scope condition (b) (“task orientation”) is closely related to the creation of the belief, specified in scope condition (a), that the task is valuable. Assuming that belief, the instructions will encourage the performers to try to do their best to achieve a correct solution to the patterns on which they will be working.

The source of evaluations identified in scope condition (c) is operationalized through the combination of the perceived scientific status of the project and the graduate-student status of the experimenter. Her age and demeanor also enhance her overall superior status relative to the participants. She will administer a “test” of contrast sensitivity in the first phase of the session and will inform the participants of their results. Note that the experimenter does not generate the scores; rather, these derive from the test. Because group members have no basis for evaluating the performances, and the setting is associated with university research, it is highly likely that they will not only accept the scores as a better indicator of ability than any judgment that they can make by themselves but also that they will consider the scores accurate. Other possible bases for performance evaluation are blocked by informing each pair of participants that they are peers in several respects (e.g., sex category, age, and year at the university).

Scope condition (d) (“collective orientation”) is implemented by the task changes that occur in the second phase: after making an initial choice, each person will receive information on the partner’s selection and will then be asked to make his or her own final choice. The instructions will emphasize that the pair should be working as a team, and that it is not only useful but also appropriate to consider and follow the other person’s choice when making a final decision.

## 2 *Independent Variables*

Levels of performance evaluation are operationalized by two clearly different scores for self and partner, one first-rate and the other poor. In the control group, no scores will be given. Levels of task diagnosticity will be created by statements by the experimenter advising participants that the 20 patterns of the test either have or have not been proven to be a very reliable indication of a person’s contrast sensitivity.

### 3 Dependent Variable

During the 25 trials of the second phase, the partner's initial choice will be programmed to show on the monitor that he or she disagrees with self on 20 trials and agrees on 5. These agreements are included because the occurrence of disagreements on *all* trials would arouse suspicions; instead, the agreements will take place every 4–6 trials in a way that does not suggest a prearranged sequence. The disagreements provide the opportunity for rejection of influence from the partner to be assessed. Thus, the high ambiguity of the task creates uncertainty in self about the correctness of a response; this, in turn, increases self's reliance on the only information that can help him or her decide on the "right" answer—namely, the scores (or their absence) during the first phase and the diagnosticity of the test. A final choice in which self stays with his or her initial choice after a disagreement is an indicator of influence rejection. The level of this variable is measured by the proportion of "self" or "stay" responses over these trials.

### 4 Irrelevant Variables of Theoretical Interest

As mentioned previously, I highlight two factors in this category: (1) level of the source's perceived superiority over self in ability to evaluate performances; and (2) level of consistency, across the evaluations, of one performer's superiority over the other. It follows from the status of "theoretically irrelevant" that I assign to these variables that they can be operationalized at any level. However, to avoid the possibility of introducing additional factors (particularly if the sample size were to be small), it is a good design feature to keep them constant *within* a given experiment. Thus, the procedures to create the difference between self and source in ability to evaluate will be the same throughout the experiment, and so too will be the sequence of correct and incorrect scores given to self and other.

Let us now turn to the other two types of particular features.

## B Truly Irrelevant Variables

My design also contains the implicit decision that a very large variety of factors (of different levels of abstractness) are truly irrelevant to the process under study. Examples of particular features of the experiment that reflect these factors are the time of the day the sessions are carried out and the type of music the experimenter likes. These features are never listed; if a researcher were to be asked explicitly about them, the answer would be that they have no actual or presumed theoretical linkage to the hypotheses.

I also count in this group some of the changes I have made in my computerized version of the standardized setting, such as the manner in which agreements and disagreements are conveyed to the participants at each trial. In this version, the monitor shows both the color of self's choice and that of the partner's "choice" and also a statement indicating either "partner agrees" or

“partner disagrees.” In the original version, self can see whether the two have either agreed or disagreed from a display of lights showing self’s initial answer and the partner’s “response.” That is, I consider the impact of the statements on the monitor to be a nonsignificant addition to what is conveyed by the lights in the original version. To my knowledge, there is no research evidence that disconfirms this idea conclusively.

## C Experimental Limitations

There are also variables that have an in-between status. They are neither theoretically irrelevant nor truly irrelevant; that is, they have not been made a part of the hypotheses. However, they *could* be of relevance, and it is for that reason they are listed in the procedures and either kept constant or within a narrow range. They are also often mentioned in the interpretation of the results as part of ad hoc explanations. I call these features “test (or experimental) limitations” (Foschi, 1980). In the present design, examples are the educational status of the participants, their age, the amount they are being paid, the sex category of the experimenter, and the country in which the experiment is being conducted. Because these factors are formulated at a low level of generality, they could not have been incorporated to the hypotheses in their current form; for that, they would have to be treated as instances of more abstract notions (Foschi, 1997, pp. 544–545). (For an interesting, compatible proposal on test limitations, see Poortinga & van de Vijver, 1987.)

Let us consider, for example, payment level. In the study I propose here, that level (as stated in [Section III](#)) is consistent with the amount currently offered at my university for taking part in social science experiments. Personal experience indicates that paying less would not be sufficient to motivate participation, whereas paying more would not significantly increase the incentive. Many students volunteer out of curiosity and because it is a common undergraduate experience to participate in experiments. Still, because level of payment *could* have an effect, it is important to mention that amount in the report on the study so the reader can arrive at his or her own conclusions about any possible effects from this factor.

In my view, the test limitation that would be most worthwhile to investigate in the present example is the sex category of the experimenter relative to that of the participants. This factor could have a role in the extent to which the experimenter and the project are accepted as a source; results from Foschi and Freeman (1991) and Foschi et al. (2001) point in that direction. My approach would be to view “sex category” as an instance of the theoretical variable “status characteristic.” I would then not only replicate the study with a male experimenter (while not changing any of the other manipulations) but also investigate other possible status factors, such as that person’s age and ethnicity. (For work on the related topic of evaluator’s status, see Webster & Sobieszek, 1974, Chapter 6.)

## VI FURTHER COMMENTS ON OPERATIONALIZATIONS

### A What Cannot and What Should Not Be Done

It is important to realize that some operationalizations are either difficult to implement or simply cannot be carried out. Thus, a theoretical hypothesis may remain experimentally untested because some or all of its variables/values cannot be manipulated in the laboratory. In some cases, this is due to practical reasons. For instance, a researcher may not be able to secure the funds (for laboratory facilities, equipment, or payment to participants) that his or her experiment requires. There are also ethical reasons why some operationalizations *should not* be done. For example, in my proposed experiment, it would be unethical to try to increase the perceived value of contrast sensitivity by informing the group members that their scores are a very reliable indication of either high or low levels of intelligence. Intelligence carries strong emotional connotations for many people, and using it as an experimental manipulation might subject such people to unacceptable levels of anxiety.

Let us focus on problems that may emerge about the feasibility of the operationalizations in relation to the choice of participant population. First, one should ensure that there are no language or other comprehension-related barriers to the understanding and acceptance of the instructions. For example, this becomes a significant design concern if the participant population contains individuals with varying levels of proficiency in the language used in those instructions. Moreover, the researcher should be aware that the participants' attitudes and beliefs could play an important role in the acceptance of the instructions. In all cases, it is essential to pretest the procedures with respondents from the same population as those who will be taking part in the experiment proper to determine whether or not the experiment is viable in all of these respects (see [Section VII](#) of this chapter, and [Rashotte, Webster, & Whitmeyer, 2005](#)). The next three examples illustrate some of these issues in relation to the standardized design:

- It could be that despite the university laboratory setting and what the researcher considers to be convincing arguments presented in the instructions, those taking part are not prepared to accept that contrast sensitivity is a valuable ability. They may also be skeptical about the value of scientific research in general. If more than a few of them fall into these categories even after strengthening the relevant instructions, it would be necessary to plan the test with either a different population or a different ability.
- The creation of a collective orientation in the laboratory may involve more than simply conveying instructions to that effect. Cross-cultural research has identified “cooperation–competition” as a useful dimension for the classification of societies and cultures (including segments within them) in terms of their central norms. “Collectivism–individualism” is a closely related concept. Although the number of studies investigating

whether these dimensions affect group-level interaction is still rather limited, there is work showing that differences at that level are in line with these dimensions (Mann, 1988, pp. 192–193; Miller-Loessi & Parker, 2003, pp. 539–542). It should then not be surprising to anticipate that an experiment that requires collective orientation would be more difficult to implement in societies that stress competitiveness than in those that place more value on cooperation. The following should also be considered: some societies may emphasize cooperation to such a high degree that task orientation is compromised; that would be the case if an answer from a higher status person is almost always accepted regardless of its merit. (On status effects in the standardized setting from either a collective or an individual orientation, see Dippong, 2012; Wilke, Young, Mulders, & de Gilder, 1995.)

- A number of expectation states experiments have investigated the linkage between status and task, and contrast sensitivity has proven to be a highly versatile instrument for this purpose. For example, this task has been credibly introduced to the participants as: (1) “masculine” (Foddy & Smithson, 1999; Foschi, 1996; Wagner, Ford, & Ford, 1986); (2) either “masculine” or “feminine” in different experimental conditions (Foddy & Smithson, 1989; Rashotte, 2006); or (3) “of no known relationship to gender” (Foschi & Lapointe, 2002). Researchers have also been successful in presenting contrast sensitivity as either explicitly or not explicitly associated with characteristics such as education level, military rank, and age (Berger, Cohen, & Zelditch, 1972; Freese & Cohen, 1973; Moore, 1968).

In the case of gender, however, results from two of the studies linking it directly to the task (Foddy & Smithson, 1999, p. 317; Foschi, 1996, p. 246) indicate that although the performers behaved as if they had accepted the instruction that the task was masculine, they were reluctant to admit this in written questionnaires. (On this point, see also Section VIII.) The lack of correspondence between behaviors and self-reports suggests that gender may have now lost some of its status value (both as a source of information and as a norm) for the participants in those two experiments.

It is also worth noting that in those two studies, the operationalization of the task’s sex linkage consisted of informing the performers that research had found men to be better than women at solving contrast sensitivity problems. Suppose that I wanted to replicate the experiment proposed here, but this time defining the ability as masculine in some experimental conditions and as feminine in others. Although the possibility of reactivity should always be taken into account, the operationalization could be made stronger by, for instance, including charts and graphs supporting the sex-linkage claim. One should also consider that it is likely (as well as hoped for) that in a not-too-distant future, participants will not readily accept that there are skills that are associated with ascribed attributes such as sex category and skin color.

## B Replications and Multiple Operationalizations

It is always useful to plan the test of a theoretical hypothesis so that the evidence originates in more than a single study. Thus, experiments need to be replicated. Following Aronson, Ellsworth, Carlsmith, and Gonzales (1990), I distinguish between two types of replications: “direct” and “systematic.” In a direct replication, a study is carried out in such a way that the researcher attempts to duplicate, as closely as possible, the procedures of the original design. This type of replication has also been called “exact” (Campbell & Stanley, 1966, pp. 32–33; Hendrick & Jones, 1972, pp. 356–357). In addition, this term is applied to the repetition of a session or “run” of an experiment that is done to obtain the desired number of respondents per condition. It is worth noting that even in direct replications, one should expect variation (either across sessions or studies) in the factors that are considered to be truly irrelevant.

In a systematic replication, a study is repeated either: (1) varying the operationalizations of theoretical concepts; or (2) letting the experimental limitations take on other values. For instance, a systematic replication of my proposed experiment could involve using a task other than contrast sensitivity, provided the task has the features specified in scope condition (a) of the hypotheses. Examples of studies incorporating such different tasks are Berger and Conner (1969), Freese and Cohen (1973), Foschi et al. (1985), and Martin and Sell (1980). With regard to the experimental limitations, I would focus on varying the characteristics of the experimenter, as I outline in Section V. Systematic replications strengthen the evidence by reducing the possibility that results are confounded with particular operationalizations. Furthermore, this type of replication can contribute to theoretical growth when experimental limitations are recast in theoretical terms. (I discuss this point in reference to cross-cultural replications in Foschi, 1980.)

## VII MANIPULATION CHECKS

Previously in this chapter, I referred to an experiment as a stage or framework within which several factors are controlled and a few others vary freely. A manipulation check is a procedure designed to verify that the controlled factors have indeed been implemented as expected. Without this check, the stability of the framework is not certain. Thus, it is an essential part of an experiment to include manipulation checks of scope conditions, independent variables, and irrelevant variables of theoretical interest, as well as of test limitations. It is equally important to report the results of these checks.

In the present experiment, the checks will be similar to those used in comparable expectation states studies. Some of these procedures will involve verifying that there has not been any experimental error (e.g., in the reading of the instructions and in the assignment of scores to the participants); this can be done by inspecting the logs that the experimenter will keep for each session. Other checks will be as follows. Levels of influence rejection obtained from the control group

can serve to establish that respondents held no prior self–other expectations. Moreover, the pattern of self-responses can be examined to determine if participants show lack of interest (e.g., by always choosing the same color as the initial response or by unvaryingly alternating between accepting and rejecting influence). Most of the checks that I plan, however, will be accomplished through the postexperimental questionnaire and interview.

It is generally understood that manipulation checks should be carried out on the scope conditions and the independent variables. Still, as mentioned previously in this text, such checks should also be done on the irrelevant variables of theoretical interest that have been identified and controlled because these procedures imply that they have been recognized as part of the hypotheses. Similarly, test limitations should be checked to verify that each of these factors had indeed either been kept constant or varied within a narrow range as intended. In this way, their inclusion in the interpretation of the results will be on firm ground.

The rest of this section outlines some of the questions I would use in the postexperimental instruments.

## A Independent Variables

Regarding level of performance, I would ask the respondents to recall the scores obtained by self and partner and to assess the level achieved by each person on bipolar 5-point scales ranging from “my partner’s performance was much better than mine” to “my partner’s performance was much worse than mine.” I would evaluate the perceived diagnosticity of the test through several questions about self’s confidence that the results from that test can be reliably generalized to assessments of task ability.

## B Scope Conditions

I would use a variety of questions in this area. I would investigate the perceived level of difficulty of the trials, and the extent to which participants were convinced that contrast sensitivity was both valuable and specific and had no prior expectations regarding self’s and partner’s task performances. In addition, I would assess the perceptions of the project and the experimenter as the source of evaluations and also the degree to which performers understood and accepted the scores they received. Finally, I would use several questions formulated as Likert scales with five categories of agreement each (from “strongly agree” to “strongly disagree”) to check on task and collective orientation. The following are two examples of these questions: “It was really too difficult to try to figure out which pattern had more red, so I just guessed” (for task orientation) and “Agreeing as a team regarding the correct decision was more important to me than my own choice” (for collective orientation). In turn, for each scope condition, my exclusion rule for the data analysis would require that a person give extreme values in the wrong direction over several questions.

## C Irrelevant Variables of Theoretical Interest

In an earlier section, I suggested that variables that have been explicitly identified as belonging to this category should be kept constant throughout the experiment. I would thus check that the source's perceived ability to evaluate performances (relative to self's) is similar across participants and that there are no major discrepancies between the received and perceived sequences of "correct" and "incorrect" scores.

## D Test Limitations

Although these features are not theoretical variables, researchers often highlight them because they *could* be in that category. It is therefore important to verify that their levels have been created as intended. As discussed previously in this chapter, I would focus on two test limitations: level of payment and sex category of the experimenter relative to the participants. I would therefore probe about the perceived fairness of the payment and the extent to which the fact that the experimenter was a woman affected the acceptance of the scores. Because the latter item concerns gender as status, it is important that questions about it be as unobtrusive as possible (see the following discussion). One possibility would be to carefully mask items about the experimenter's sex category with items about other characteristics, both status and nonstatus, of this person.

# VIII FURTHER COMMENTS ON MANIPULATION CHECKS

## A Exclusions

A key issue related to manipulation checks concerns the exclusion of participants from the data analysis. The topic is often misunderstood. Because the test of a hypothesis assumes that the scope conditions and the independent variables have been implemented as intended, it is legitimate to exclude from the data analysis those for whom that implementation has failed.<sup>9</sup> Rejection rules, however, should be explicitly formulated beforehand to avoid the possibility of the exclusions being misunderstood as an attempt to keep only those data that support the researcher's hypotheses. The rules should also be as conservative as possible because a substantial proportion of exclusions (often taken to be 20% or higher of the total number of participants) opens the door to compromising random assignment. If an experiment results in that level of exclusions, it is wise to rethink the design and the procedures and to redo the study.

---

9. Because, as mentioned previously, it is only for design reasons that irrelevant variables of theoretical importance are kept constant, the consequences of failing manipulation checks in their case does not warrant exclusions. Similarly, if a test limitation fails a manipulation check, this does not affect the test of the theoretical hypotheses—only the post hoc interpretation of the results with respect to the factor in question.

## B Behaviors and Self-Reports

As described previously here, most manipulation checks in the standardized setting utilize self-reports obtained through the postexperimental questionnaire and interview. It is important to remember that self-reports are a person's reflection of what has occurred during the session and may thus be affected by factors such as memory, self-presentation norms regarding one's task ability, level of support for the perceived goals of the research project, and even suspicion. Self-reports also increase a respondent's awareness of his or her answers, and this is particularly a concern if the topic is of a sensitive nature. The behavioral measure of rejection of influence, on the other hand, is likely to be less obtrusive and hence to avoid these effects. Yet, self-reports are valuable tools in assessing manipulation checks because often it is not feasible to obtain behavioral data on all variables of interest. Every effort should be made to control for factors that may affect these checks; at the very least, one should be apprised of the possible effects of these factors when interpreting the results. (The relationship between self-reports and behaviors is a classic topic in social science methodology and has been examined by many authors; for a useful discussion of this topic in the context of expectation states research, see [Driskell & Mullen, 1988](#).)

As an example of the different responses often obtained from the two types of measures, let us consider the results from the [Foschi et al. \(2001\)](#) experiment mentioned in [Section III](#). Each same-sex pair received scores indicating that one person clearly had more ability than the other. Rejection of influence showed no effects from sex category of dyad (as predicted, because participants were peers in this respect). However, there were significant effects from sex category of participant in the self-reports: at every level of scores received (in each case, the same for male and female dyads), men reported higher levels of ability relative to the partner than did women. The authors interpreted this finding in terms of the operation of different, gender-based self-presentation norms that emphasize modesty in women but not in men.

The following is a useful procedure that minimizes the effects of extraneous variables on self-reports. It consists of dividing prospective participants at random into two groups—one to be assigned to the experiment proper and the other to provide information on those variables that require manipulation checks. An often-cited example of this procedure is [Goldberg's 1968](#) study. In the experiment proper, the extent to which the respondents were biased by the sex category of a (fictitious) performer was measured; the other group provided data on their perceived sex linkage of various occupations. In this way, this second group supplied information that was not affected by possible bias in the task of assessing performances. Moreover, because of random assignment, results on sex linkage of occupation from the latter can be assumed to apply to a similar extent to those participating in the experiment proper.

Finally, the inclusion of control groups as part of a standardized expectation states experiment provides an effective way of doing manipulation checks that yield behavioral data. [Wagner et al. \(1986\)](#) employed such groups

to assess the combined effects from gender-as-status and the instructions that contrast sensitivity was a valuable, masculine task. Similarly, Foschi and Lapointe (2002) used such control groups to assess the extent to which gender functioned as a status characteristic for those in their study. In both cases, the control groups avoid asking performers directly about possibly sensitive issues.

## IX SUMMARY AND CONCLUSIONS

In this chapter, I examined several issues concerning the experimental test of theoretical hypotheses. I began by identifying the different types of factors involved in such a test: independent, intervening, and dependent variables; scope conditions; test limitations; irrelevant variables of theoretical interest; and truly irrelevant variables. I focused on operationalizations and manipulation checks in the context of laboratory experiments. Operationalizations are a researcher's translation of theoretical variables into procedures that render them observable. In turn, the effectiveness of these procedures must be checked. I proposed an example involving a set of hypotheses within the expectation states tradition and an experimental design to test for them, and I used this example to illustrate a variety of methodological topics. Thus, I discussed limits on operationalizations in terms of what cannot and should not be done, and I emphasized the value of replications and multiple indicators.

Regarding manipulation checks, I examined the reasons why the exclusion of some data may be necessary and compared behaviors and self-reports on the type of information that they yield. Carefully designed and executed laboratory experiments are an invaluable instrument in the construction and test of theories. Throughout the chapter, I offered ideas on how to enhance the design and interpretation of such experiments.

## ACKNOWLEDGMENTS

This is an updated version of the chapter that appeared in the first edition of *Laboratory Experiments in the Social Sciences* (2007). Preparation of the original chapter was supported by Standard Research Grant No. 410-2002-0038 from the Social Sciences and Humanities Research Council of Canada. I gratefully acknowledge this support. I also thank Lok See Loretta Ho, Vanessa Lapointe, and Maria Zeldis for valuable comments.

## REFERENCES

- Aronson, E., Ellsworth, P. C., Carlsmith, J. M., & Gonzales, M. H. (1990). *Methods of research in social psychology* (2nd ed.). New York: McGraw-Hill.
- Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51, 1173–1182.

- Berger, J., Cohen, B. P., Zelditch, M., Jr. (1972). Status characteristics and social interaction. *American Sociological Review, 37*, 241–255.
- Berger, J., & Conner, T. L. (1969). Performance expectations and behavior in small groups. *Acta Sociologica, 4*, 186–198.
- Berger, J., Fisek, M. H., Norman, R. Z., Zelditch, M., Jr. (1977). *Status characteristics and social interaction: An expectation-states approach*. New York: Elsevier.
- Berger, J., Webster, M., Jr. (2006). Expectations, status, and behavior. In P. J. Burke (Ed.), *Contemporary social psychological theories* (pp. 268–300). Stanford, CA: Stanford University Press.
- Bianchi, A. J. (2010). Status characteristics theory/expectation states theory. In J. M. Levine, & M. A. Hogg (Eds.), *Encyclopedia of group processes and intergroup relations* (pp. 845–849). Thousand Oaks, CA: Sage.
- Camilleri, S. F., & Berger, J. (1967). Decision-making and social influence: A model and an experimental test. *Sociometry, 30*, 365–378.
- Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.
- Cohen, B. P. (1989). *Developing sociological knowledge: Theory and method* (2nd ed.). Chicago: Nelson-Hall.
- Correll, S. J., & Ridgeway, C. L. (2003). Expectations states theory. In J. Delamater (Ed.), *Handbook of social psychology* (pp. 29–51). New York: Kluwer/Plenum.
- Dippong, J. (2012). The effects of scope condition-based participant exclusion on experimental outcomes in expectation states research: A meta-analysis. *Social Science Research, 41*, 359–371.
- Driskell, J. E., Jr., & Mullen, B. (1988). Expectations and actions. In M. Webster Jr., & M. Foschi (Eds.), *Status generalization: New theory and research* (pp. 399–429). Stanford, CA: Stanford University Press, 516–519.
- Foddy, M., & Smithson, M. (1989). Fuzzy set and double standards: Modeling the process of ability inference. In J. Berger, M. Zelditch Jr., & B. Anderson (Eds.), *Sociological theories in progress: New formulations* (pp. 73–99). Newbury Park, CA: Sage.
- Foddy, M., & Smithson, M. (1999). Can gender inequalities be eliminated? *Social Psychology Quarterly, 62*, 307–324.
- Foschi, M. (1980). Theory, experimentation, and cross-cultural comparisons in social psychology. *Canadian Journal of Sociology, 5*, 91–102.
- Foschi, M. (1996). Double standards in the evaluation of men and women. *Social Psychology Quarterly, 59*, 237–254.
- Foschi, M. (1997). On scope conditions. *Small Group Research, 28*, 535–555.
- Foschi, M., Enns, S., & Lapointe, V. (2001). Processing performance evaluations in homogeneous task groups. *Advances in Group Processes: A Research Annual, 18*, 185–216.
- Foschi, M., & Freeman, S. (1991). Inferior performance, standards, and influence in same-sex dyads. *Canadian Journal of Behavioural Science, 23*, 99–113.
- Foschi, M., & Lapointe, V. (2002). On conditional hypotheses and gender as a status characteristic. *Social Psychology Quarterly, 65*, 146–162.
- Foschi, M., Warriner, G. K., & Hart, S. D. (1985). Standards, expectations, and interpersonal influence. *Social Psychology Quarterly, 48*, 108–117.
- Freese, L., & Cohen, B. P. (1973). Eliminating status generalization. *Sociometry, 36*, 177–193.
- Goldberg, P. (1968). Are women prejudiced against women? *Transaction, 5*, 28–30.
- Hendrick, C., & Jones, R. A. (1972). *The nature of theory and research in social psychology*. New York: Academic Press.
- Jasso, G. (1988). Principles of theoretical analysis. *Sociological Theory, 6*, 1–20.

- Lovaglia, M. J., & Houser, J. A. (1996). Emotional reactions to status in groups. *American Sociological Review, 61*, 867–883.
- Magnusson, E., & Marecek, J. (2012). *Gender and culture in psychology: Theories and practices*. Cambridge, UK: Cambridge University Press.
- Mann, L. (1988). Cultural influences on group processes. In M. H. Bond (Ed.), *The cross-cultural challenge to social psychology* (pp. 182–195). Newbury Park, CA: Sage.
- Martin, M. W., & Sell, J. (1980). The marginal utility of information: Its effects upon decision-making. *Sociological Quarterly, 21*, 233–242.
- Miller-Loessi, K., & Parker, J. N. (2003). Cross-cultural social psychology. In J. Delamater (Ed.), *Handbook of social psychology* (pp. 529–553). New York: Kluwer/Plenum.
- Moore, J. C., Jr. (1968). Status and influence in small group interaction. *Sociometry, 31*, 47–63.
- Poortinga, Y. H., & van de Vijver, J. R. (1987). Explaining cross-cultural differences: Bias analysis and beyond. *Journal of Cross-Cultural Psychology, 18*, 259–282.
- Rashotte, L. S. (2006, July). Controlling and transferring status effects of gender. In *Paper presented at the annual meeting of the International Society of Political Psychology, Barcelona*.
- Rashotte, L. S., Webster, M., Jr., & Whitmeyer, J. M. (2005). Pretesting experimental instructions. *Sociological Methodology, 35*, 163–187.
- Wagner, D. G. (1984). *The growth of sociological theories*. Beverly Hills, CA: Sage.
- Wagner, D. G., Ford, R. S., & Ford, T. W. (1986). Can gender inequalities be reduced? *American Journal of Sociology, 51*, 47–61.
- Webster, M., Jr., & Sobieszek, B. (1974). *Sources of self-evaluation: A formal theory of significant others and social influence*. New York: Wiley.
- Wilke, H., Young, H., Mulders, I., & de Gilder, D. (1995). Acceptance of influence in task groups. *Social Psychology Quarterly, 58*, 312–320.
- Zelditch, M., Jr. (2013). Thirty years of advances in group processes: A review essay. *Advances in Group Processes: A Research Annual, 30*, 1–19.

## Chapter 12

# The Standardized Experimental Situation in Expectation States Research

## Notes on History, Uses, and Special Features

Joseph Berger

*Stanford University, Stanford, California*

### I INTRODUCTION

Expectation states theory is a *theoretical research program*. As such, it consists of a set of interrelated theories, bodies of relevant research concerned with testing these theories, and bodies of research that use these theories in social applications and interventions (Berger, 1974).

A major substantive concern of these theories has been with understanding how status processes operate to organize interaction in groups: the different conditions under which these processes are activated; the various forms these processes can assume; the stable states that emerge as status processes evolve over time; and the different types of behaviors determined by these stable status states.

In undertaking the study of status and other social processes within the expectation states program, we have been guided by a specific theory building strategy. The first principle of our strategy has been to analytically isolate status processes from other processes with which they may be interrelated in some concrete setting, such as social control, affect, or power processes. The idea here is that the interrelation of different interpersonal processes is to be treated as a task for subsequent theoretical research. A second key principle in our strategy has been to describe the operation of status and other processes in terms of abstract theories that are formally rigorous and empirically testable. Finally, a third key principle of our strategy has been to develop theory-based empirical models in order to be able to apply our abstract theories to the realities of specific concrete social settings (see the appendix at the end of this chapter).

Testing the type of abstract theories we have been interested in constructing often involves creating special status conditions, as for example, conditions in which status processes actually are separated from affect or control processes or creating social conditions that are rarely found in everyday social interaction. As a consequence, our theoretical strategy led us to the conviction that a major (although not exclusive) source of data for this program would have to be from experimental research. It is within this context that the standardized experimental situation (SES) has evolved as an important source of data for testing and developing theories in the expectation states program.

The major elements of this standardized experimental setting were devised in the late 1950s and early 1960s, although research on the nature and properties of this setting continues to the present. In this chapter, I briefly describe the early history in the construction of this experimental situation as well as some of its special features. In addition, I consider some of the ways the experimental situation has been used to both develop and apply the theoretical knowledge of the different branches of the expectation states program.

Throughout, I describe events that date back 40–45 years and have not been previously recorded as history. Thus, as is true of all such historical narratives, there are cautions that the reader should bear in mind in reviewing this history. Two are particularly important. I undoubtedly present a much more simplified picture of events than in all likelihood was true, and I present a more rational picture of our activities than probably was the case. Actually, there was a considerable amount of trial and error in our activities, which is an important fact to bear in mind in understanding the evolution of the standardized experimental situation.

## II ON THE CONSTRUCTION OF SES

The history of SES begins with my work following the completion of my Ph.D. thesis ([Berger, 1958](#)). The general concern of the thesis was with accounting for the properties of interaction hierarchies as they emerged in small problem-solving groups. These properties were most evident in Bales' observations of small, informal task groups whose members were presumably similar in status (see [Bales, 1953](#); [Bales & Slater, 1955](#); [Bales, Stodbeck, Mills, & Roseborough, 1951](#); [Heinecke & Bales, 1953](#)). Bales found that inequalities in initiation of activity, in receipt of activities, on ratings of best ideas, and in group guidance regularly emerged in such groups. Once they emerged, these inequalities tended to be stable, and with the possible exception of sociometric rankings they tended to be highly intercorrelated. Research by others (e.g., [Harvey 1953](#); [Sherif, White, & Harvey, 1955](#); [Whyte, 1943](#)) had shown that the evaluations of specific performances of individuals were correlated with the individual's position in the established hierarchy of the group. Independent of actual performance level, high-status members are seen as performing better than low-status members.

I approached the problem of explaining the nature of such interaction hierarchies by conceptualizing an idealized interaction process as it occurs in status-homogeneous groups. This process was seen as involving sequences of behaviors such as action opportunities (chances to perform), performance outputs (problem-solving attempts), performance evaluations, and communicated evaluations (positive and negative reward reactions). Basically, I argued that expectations (or expectation sets as they were then called) *emerged* out of the evaluations of performances and created a status ordering that determined the distribution of subsequent behaviors. Furthermore, given my arguments on the way in which this idealized interaction process operates, I sought to account for the *interrelation* of distributions of action opportunities, performance outputs and rewards to overall performance evaluations, as well as account for the *stability* of this structure across tasks given this interrelation. In addition, I argued that all of these processes would be affected by the primary normative orientation under which the group operated. Specifically, in “instrumental” (or task) focused groups, the behavioral components of the status hierarchy would be more highly interrelated and the structure more stable than in “integrative” (or process) oriented groups.

At this stage, I believed that I had only a general conception of what appeared to be fundamental processes. I wanted to expand the scope of these ideas and at the same time get a much more precise understanding of what was involved in their operation. For example, in the initial formulation expectations emerged out of the evaluation of performances. Were there other processes by which expectations emerged or by which they became significant in a situation of action? Also, in my thinking I had developed a conception of an underlying process (an expectation process) that was related to an observable process that, with the addition of influence behaviors, became known later as the *observable power and prestige order* (OPPO). An important unanswered question was, “How can we conceptualize the different ways in which these two kinds of processes are related?”

Eventually, following the dissertation, I came to believe that in order to study these processes with the precision that I was interested in, I had to be able to create performance expectations. Specifically, I wanted to study the behavioral consequences of high or low performance expectations possessed by an actor and his interacting partner. Because I wanted to be able to *randomly assign* such performance expectations to self and other, it became clear that the techniques required to create these conditions would have to involve the manipulation of information given to subjects. Subsequently, I also realized that I had to devise a more highly controlled experimental situation in which to study the behavior of expectation state processes.

## A The Manipulation of Expectations

A solution to the first problem turned out to be not too difficult. I decided that I had to create an ability or set of abilities that is instrumental to solving a group

task. This would have to be one that subjects would believe was an actual (real) ability and one about which they had no idea as to their actual capabilities. Inasmuch as the group problems we most frequently used in those days involved human relations problems, the answer was clear: social insight and prediction ability. This was defined to subjects as the ability to “get right into a social situation,” and “understand what’s going on,” and “to predict quite accurately what will happen next.”

In 1957 and 1958, while at Dartmouth College, a colleague<sup>1</sup> and I devised and ran a two-phase pilot study that involved two-person problem-solving groups. In the first phase, subjects were informed about the discovery of social insight and prediction ability. They were then “tested” on this ability and were randomly assigned high and low scores. Thus, we created a situation in which one individual in each group was in a high-self and low-other state and the second was in a low-self and high-other state. In the second phase, subjects were asked to solve in an open-interaction setting a complex human relations case that involved using their social insight and prediction ability. In addition, subjects were presented with individual and group standards, purportedly based on studies of college students like themselves, to enable them to evaluate their individual as well as group performance. Who-to-whom interactions were scored using the behavior categories formulated in the thesis, and we predicted that the distribution of these behaviors would be a function of the performance expectations we had created.

We found that subjects differed on the distribution of action opportunities, with the low-status subject distributing more to the high than the high to the low; that they differed on performance outputs, with the high initiating more than the low; and that they differed on positive rewards, with the low distributing more to the high than the high to the low. Viewing these profiles from the standpoint of *performer-reactor differences* (problem-solving attempts initiated to distribution of action opportunities and rewards to the other), this meant that performer-reactor rates were higher for the high-status subject compared to his low-status partner.

We were impressed with the effectiveness of these techniques, and I considered that this problem was solved—that we could, using such methods, manipulate performance expectations and study their consequences.

## B The Behavioral Setting

At this point, I still did not have the kind of experimental situation that I believed would enable us to study expectation processes with the precision that I desired. Actually, I was getting ready to do a full-scale open-interaction experiment based on what had been discovered in the Dartmouth pilot study. This experiment was intended to investigate the behavioral consequences of different types of expectation structures (Berger, 1960).

---

1. This test study was done in collaboration with the social psychologist Harry A. Burdick.

For a time, we considered developing a modified Bavelas Box.<sup>2</sup> This would enable us not only to study the behavioral consequences of expectation states but also to control the rates at which action opportunities, performance outputs, rewards, and exercised influence occurred between interactants. The assumption was that differences in the rates of each of these behaviors would lead to the formation of different types of expectation states. In fact, I remember exploring this idea in the late 1950s with my colleague and future collaborator, Morris Zelditch.

However, just as I was leaving Dartmouth and preparing to join the Department of Sociology at Stanford University, I started to work on a highly precise formulation of how expectations determine behavior and how behavior determines expectations. This was a finite Markov chain model developed in collaboration with J. Laurie Snell, a mathematician at Dartmouth College (Berger & Snell, 1961). The key behavior in this model was *influence behavior* that, from my standpoint, involved *changes in the performance outputs of an actor as a consequence of the behavior of another*. I realized that in order to study these influence behaviors, which by this time were conceptualized as part of the observable power and prestige order, I needed an interaction sequence that would involve a large number of disagreements between interactants, and so the decision-making structure of the standardized experimental situation was born. I immediately went to work on more fully conceptualizing this decision-making situation and particularly the social context within which it should take place—such as task and collective orientation. Fortunately, I had a graduate student working with me, Robert Z. Muzzy, who informed me that he could build a machine that would do what I wanted it to do—control the behavior of interactants in the ways that I wanted it controlled—and once I described the decision-making process to him, he built our first interaction control machine (ICOM). The ICOM allowed interaction between participants but eliminated the face-to-face component so that the only information known to subjects was that provided by the experimenter.

When finalized, the basic decision-making sequence had the following features: each individual—for example, in a two-person group—is given on every trial 5–10 seconds to study a problem and decide which one of two alternatives is the correct solution to the problem. His or her choices are then communicated to the partner. The communicated information on choices is controlled by ICOM so that subjects can be told on any trial whether they agreed or disagreed with their partner independent of what is actually the case. With this feedback information, each individual then has the opportunity to restudy the problem and to make a final decision. These final decisions are not communicated to partners. Initial choices are defined as only preliminary decisions to provide

---

2. In 1950, Alex Bavelas formulated a set of concepts to determine the relative centrality of positions in communication networks (Bavelas, 1950). The Bavelas Box was a device used to create and study group performances in the networks of interest to Bavelas and his students (see Leavitt, 1950).

information to partners, while subjects are told that final decisions are the only decisions that count on their team's performance record.

Although there has been experimentation on the number of critical (disagreement) trials used in experiments, typically at least 20 critical trials have been used with either 3 or 5 neutral (agreement) trials to allay suspicion. The basic behavioral quantity obtained in this setting is  $P(s)$ , the proportion of times individuals stay with an initial decision given disagreement with their partner.<sup>3</sup>

Two major problems still remained to be dealt with before we could actually do experiments. We had to develop tasks that were appropriate to the new experimental setting, and we had to construct standardized scenarios that would define the experimental situation to subjects in a meaningful manner.

## C Experimental Tasks

Developing an appropriate task for the new experimental situation presented problems. Initially, we considered using the social insight and prediction ability task that had been so successful in the Dartmouth research. We actually constructed such a task involving 20–30 short decision-making situations, but it proved to be unsuitable. There was simply too much information in each of these cases for subjects to fully process within the short time span of a 5- to 10-second experimental trial. So we kept searching for a new task—a new ability.

At approximately this time, I learned of the research that members of the anthropology department were doing on the structure of a particular primitive language. This gave me the idea for constructing the “meaning insight” ability task. This task involved matching English words with phonetically presented words from a primitive language that presumably had the same meanings as the English words. On each problem trial, subjects were told to study the English words and associate whatever meanings they called to mind. They were then told to study the non-English words, sound them to themselves, and associate whatever meanings they called to mind. Their task was to decide which primitive words had the same meaning as the English word.

My students and I constructed a large number of meaning insight items in two forms. In the first, two primitive words and one English word were used, and in the second two English words and one primitive word were used. The primitive words, of course, were fictional. We administered these items to students in classroom settings and selected as usable those items in which the choice alternatives were equally likely to be chosen or as close to that criterion as we could get.

---

3. It was assumed that it would be possible to modify this decision-making sequence to capture some of the other types of behaviors in the OPPO. For an example of just such a modification in decision-making sequence, see the study by [Conner \(1977\)](#), in which the dependent variable is the likelihood of making performance outputs given action opportunities.

Despite some problems connected with it, we decided that meaning insight was a usable task.<sup>4</sup> Subjects accepted the idea that such an ability actually exists. They were interested in being tested for the ability, and they believed that they “could work on the task”—in the sense that there was cognitive activity involved in getting the right solution. Thomas L. Conner did some of the most important work on this task and other tasks that were being developed in our laboratory at that time (Conner, 1964). Still later, in collaboration with Bernard P. Cohen, Morris Zelditch, and graduate students, we worked on other tasks, particularly the different forms of the spatial insight task. James C. Moore played a major role in the development of this task (Moore, 1965).<sup>5</sup>

## D Standardized Scenarios

One major task remained, namely to develop a standardized scenario for experiments in this new setting. This involved constructing appropriate experimental procedures.

We wanted to develop a set of standardized procedures that could be used in a large number of experiments. These experiments would, of course, differ in terms of the theoretically relevant conditions and variables involved. But the idea was to have a *core scenario* that would be used from experiment to experiment.

Basically, we wanted to define the experimental situation as socially meaningful and as embodying the information and features we regarded as important in doing expectation states research. Among other things, this required the following:

- Subjects should accept the reality of the instrumental characteristic, meaning insight ability. This meant accepting the idea that it was a newly discovered and very important ability, that task success or failure depended on their level of this ability and that it was not simply a result of guesswork or chance.

4. As a condition on the construction of tasks, it was believed that the choice of alternatives from one trial to a second should be an independent trials process. For example, the choice of an alternative, A or B, on trial n should be independent of the particular choice, A or B, that had been made on the n-1 trial. This condition did not hold for the first three tasks that we developed, including that of meaning insight (Conner, 1964).

5. “Spatial insight” tasks were constructed over a number of years and took different forms. Initially, each task involved a slide of a rectangular figure where the rectangle consisted of 100 subrectangles. By a random process, 50 of the subrectangles were selected to be colored black and 50 white. The subject’s task was to determine whether there were more white or black in any given rectangle. To counter a bias toward a white response that was discovered in this first task, a second version was developed in which each task consisted of two rectangles, each with different arrangements of black and white subareas. In this case, the subject’s task was to determine which rectangle had a greater amount of white area. Near-veridical versions of this task were also constructed in which a response bias of a given magnitude favoring either white or black was built into the task (Conner, 1965). See Moore (1965) for a report of the development of this task.

- Subjects should accept the idea that meaning insight ability was unrelated to other well-known abilities, such as mathematical and verbal ability. The objective was to dissociate the instrumental ability from those that subjects would have knowledge about. This would facilitate the random assignment of the different levels of meaning insight as required for experimental study.

At this stage, the notion of scope conditions, as abstractly defined conditions under which a social process is predicated to hold, was still being developed.<sup>6</sup> Already, however, it was believed that the next two conditions were essential for the study of expectation states processes:

- Subjects should be task oriented. This meant that their primary focus was on the task and that they would be motivated to work for the correct solutions of the task.
- Subjects should be team and collectively oriented. This meant that taking into account the judgments of the other person on the team was both legitimate and useful to the team goal. Subjects' initial decisions were defined as preliminary choices, and they were informed that only final choices mattered. To create task and collective orientation, we made extensive use of performance standards, both individual and group, which were presumably based on studies of subjects like themselves, here and elsewhere.

The number of disagreements (or critical trials) was an important issue:

- It was necessary to provide subjects with enough critical trials so as to get stable measures of the subject's behavior, while at the same time reducing the subject's suspicion, due to the number of critical trials, as much as was possible.<sup>7</sup>

In 1962 and 1963, we ran our first status experiment in the standardized experimental situation. In this first experiment, we manipulated expectations on C\*, the characteristic instrumental to the task, assigning individuals to high-low, low-high, high-high, and low-low self-other states. We predicted and found that the individual's likelihood of resisting influence was directly related to the individual's expectation advantage over his or her partner (Berger & Conner, 1966, 1969).

---

6. For discussions of the notion of theoretically defined scope conditions, see Berger (1974) and Walker and Cohen (1985).

7. In the earliest stages of our work, we anticipated finding trend effects in subjects' response data. Consequently, we were interested in employing enough trials so as to detect these trends and to allow the process for each subject to reach a stable state. However, after the Air Force study (Berger, Cohen, & Zelditch, 1972), which involved an extended trials sequence, we became aware of the problems—such as boredom, weakening of task, and collective orientation—associated with such extended sequences; almost all subsequent studies at Stanford involved 23–25 trials, including 3–5 neutral trials.

As we conceptualized our task back in the early 1960s, all four components—ways to manipulate or identify status and expectation states, a core set of standardized scenarios and experimental procedures, a set of novel experimental tasks, and an appropriate decision-making process—were required to construct a standardized experimental situation. Developing this situation required a considerable investment of time and effort. We accepted these costs because we anticipated that there would be a long-term payoff for our research.

### III THE GRAPH FORMULATION OF STATUS CHARACTERISTIC THEORY

The graph formulation of the status characteristic theory was introduced in expectations state research in 1977 ([Berger, Fisek, Norman, & Zelditch, 1977](#)). The overall three-part structure of the theory, its metatheoretical component, its theoretical component, and its theory-based models component are briefly described in the appendix at the end of this chapter. With the introduction of the graph formulation, the relation between theory and experimental situation was still further developed. There are at least three ways in which this occurred that are particularly worth noting.

First, the graph theory allowed us to represent in a formal manner the different types of actor-situational structures in which we were interested. From the early 1960s, we had been using “heuristic graphs” to map status situations that we had determined from experience could be created and realized in the standardized experimental situation. The new theory allowed us to transform these heuristic graphs into mathematical structures. This enabled us to deal with an extremely wide range of status situations from the simplest to the most complex. We could now formally represent these situations and make general as well as specific predictions about what could be expected to occur in these status situations.

Second, the new theory facilitated the task of devising experiments to test theoretical predictions. The status graphs have been used to conceptualize in terms of their graphic structures experimental conditions required in such tests. In addition, the theory has been used to assess the feasibility of such experimental tests. The graph theory has four parameters—*two general and two situational parameters*—whose values have to be determined to make specific numerical predictions. The general ones are assumed to be applicable across a broad range of situations and are either empirically determined, as in [Berger et al. \(1977\)](#) and [Balkwell \(1991\)](#), or determined on *a priori* grounds, as in [Fisek, Norman, and Nelson-Kilger \(1992\)](#). The two situational parameters are applicable to and have to be determined for the specific situations that are of immediate concern. Given our experience in working with the standardized situation, we have been able in general to assess the range of empirical values that we might expect to find for each of our two situational parameters. Using this information, we then

can determine the range of empirical results that might obtain for given experimental conditions. On the basis of such analysis, we have been able in general to decide whether the likely results of a proposed experiment will enable us to make a decision or not on the theoretical arguments that were being tested (see Section IV,B).

Third, on a more general level, the graph theory enables us to specify more fully what is involved in applying such a theory to a specific empirical situation. The graph theory, as is true of the original version and the second version of the status characteristic theory (Berger, Cohen, & Zelditch, 1966; Berger & Fisek, 1974), is an abstract theoretical structure. Such an abstract theory must be empirically interpreted. This is the role of what we have called *theory-based empirical models*. These models consist of a combination of factual information and assumptions—for example, factual information (or assumptions) that race or gender can be treated as instances of a diffuse status characteristic for a given population at a given time, or model assumptions on the translation of theoretical concepts to behavioral observations (see the appendix). Understanding the nature of and the specific elements in such models is essential inasmuch as such elements are involved in testing and applying an abstract theoretical structure to some concrete reality. Furthermore, it is the status of these elements that often is initially called into question with the failure of theoretical tests and applications.

## IV USES AND SPECIAL FEATURES OF SES

### A Comparability and Cumulativeness

One of the primary objectives in developing a standardized situation was to facilitate the acquisition of a cumulative body of comparable data to be used in theoretical research. Instead of being confronted with the results of a set of experiments—all presumably dealing with the same problem—in which tasks differed, settings differed, and dependent variables differed, the objective was to have such factors similar with variations being restricted to theoretically relevant factors and variables. Presumably, the results of such experiments would be particularly useful in developing, assessing, and testing theoretical knowledge. In fact, there are important situations in which this has turned out to be the case.

When my colleagues, Hamit Fisek and Robert Norman, and I constructed the graph formulation of the status characteristics theory, we were able to assess the consistency of that formulation against the results of 12 experiments (involving 57 experimental conditions) that had already been conducted in the SES (Berger et al., 1977). The fact that the formulation was consistent with this extensive database was of enormous importance in allowing us to evaluate our work at that stage, in informing us that we were on the right track, and in enabling us to build further on our theory (see also Fisek et al., 1992).

## B Strong Tests of Theoretical Arguments

It was assumed that the standardized experimental situation would be used to test and develop theoretical arguments across a wide range of substantive problems, and this in fact has been the case (see [Section IV,D](#)). Among such experiments, we were particularly interested in constructing strong tests of theoretical arguments.

*Strong tests of theoretical arguments* are empirical studies that involve: (1) direct tests of specific theoretical arguments in the context of alternative arguments; where (2) the original arguments and their alternatives are formulated within a common theoretical language. Such tests, when they are experiments, may consist of a complex set of experimental manipulations and conditions, and they may also involve a subtle pattern of predicted experimental results.<sup>8</sup>

Strong tests have been conducted successfully in the standardized situation. Three such experiments have been the [Wagner, Ford, and Ford \(1986\)](#) study on the confirmation and disconfirmation of gender-based status expectations; the [Norman, Smith, and Berger \(1988\)](#) and [Berger, Norman, Balkwell, and Smith \(1992\)](#) studies on status inconsistency and status organizing principles; and the [Webster, Whitmeyer, and Rashotte \(2004\)](#) study on second-order expectations. In each of these cases, the experiments were conducted to test for derived theoretical predictions while at the same time discriminating among alternative theoretical arguments.

In the case of the status inconsistency studies ([Berger et al., 1992](#)), for example, the experiments involved a direct test of the *principle of organized subsets*. This principle argues that *all valued status information* that has become salient is combined by an actor so as to take into account its sign (whether it is positive or negative status information) and its degree of task relevance (its weight) in the situation. Other theoretically reasonable principles have been proposed, such as *status canceling* principles, in which actors are seen to “cancel” oppositely signed and equally weighted items of status information, and *status balancing* principles, in which actors are seen as eliminating inconsistent status information in forming expectations for each other. Our objective was to design an experiment in which outcomes that are supportive of one of these principles are at the same time inconsistent with those that would be supportive of the other principles. To do this, it first was necessary to translate each of the nonstatus characteristic principles—balancing and canceling—into the concepts of the status characteristics theory and then to determine the conditions under which these principles made alternative predictions. In translating all status processing alternatives into the same theoretical language, we were creating multiple versions of the status characteristic theory (or *theoretical variants* in the

---

8. For related ideas, see [Platt \(1964\)](#) and [Cole \(2001\)](#).

language of [Wagner and Berger \(1985\)](#) and [Berger and Zelditch \(1993\)](#)). As a consequence, we were assured that the *only differences* involved in deriving alternative predictions were those due to the alternative status-organizing principles in these theoretical variants.

On the basis of extended experience in working with the SES, we had information on the range of empirical values that we might expect to find for each of the situational parameters of the graph theory. With this information, we could then determine the possible range of empirical results that might be obtained under the operation of each principle. Based on this analysis, we determined that there was a good chance of getting results that would enable us to discriminate between the different status principles. Eventually, the experiment was conducted, and the results provided support for the principle of organized subsets while at the same time they failed to provide support for the competing status-processing principles ([Berger et al., 1992](#); [Norman et al., 1988](#)).

## C A Holy Triangle: Tests, Theories, and Applications

Very early in my work with colleagues B. P. Cohen and M. Zelditch on the status characteristics theory, the issue arose of how one relates research in a highly controlled setting such as the SES to research in less controlled settings. For example, how does one generalize the results of experiments carried out in the SES to nonexperimental settings? In part, in response to such questions we started to talk about what we called within our group a “holy triangle”: tests, theories, and applications. The basic idea is straightforward. In situations in which it is required, highly controlled experimental settings are used to test theories, and it is theories and theoretical arguments that are applied to less controlled settings or used in social applications and interventions. This, of course, is an ideal conception of the development and use of theoretical ideas, and it is not clear to me how often in fact it has been realized. However, one very important case in which this ideal conception appears to have been realized is worth describing in this context.

When Hamit Fisek and I published the second version of the status characteristics theory ([Berger & Fisek, 1974](#)), we presented theoretical arguments that claimed that all of the status information that has become salient in a situation of action is combined. At that time, we had the results of experiments that we had done in SES, starting with experimental findings published in 1970 ([Berger & Fisek, 1970](#); [Berger, Fisek, & Crosbie, 1970](#)) that supported this argument. In addition, there was the as yet unpublished research of Zelditch, Lauderdale, and Stublarec that we knew about, which also supported the combining argument ([Zelditch et al., 1975, 1980](#)).

In 1977, Susan Rosenholtz, a student of Elizabeth Cohen, used the combining argument that appeared in the second version of the status characteristics theory to predict that groups exposed to multi-ability tasks would have less differentiated hierarchies than those working on standard unidimensional tasks. This research was done in an open-interaction laboratory setting and, indeed,

Rosenholtz found what she had predicted: less differentiated hierarchies in groups exposed to multi-ability tasks (Rosenholtz, 1977).<sup>9</sup> Eventually, Cohen and colleagues devised multi-ability status interventions that are suitable for classroom use, and they have been able to study their effects as part of their “complex instruction curriculum” in actual classroom settings (see Cohen, 1982; Cohen & Lotan, 1995b, 1997b).

This, then, is a case in which researchers have gone from experiments in SES on combining status information to general theoretical arguments on combining, to predictions in open-interaction laboratory settings based on these arguments, to field studies, and to interventions in actual ongoing classroom situations. A case similar to that discussed previously is to be found in the research of Murray Webster on source theory. Webster first formulated source theory (Webster, 1969), then tested the theory in SES, and subsequently he and Doris Entwistle devised interventions based on the theory that they used to raise children’s expectations in actual classroom settings (Webster & Entwistle, 1974). In both cases, research in SES was either the basis of theoretical arguments or used to test such arguments. Theoretical arguments, in turn, were then used as the grounds for applications and interventions.<sup>10</sup>

## D Flexibility of the Standardized Experimental Situation

The standardized experimental situation has proven to be a highly flexible setting in the sense that it has proven to be possible to adapt and employ the situation in the study of different theoretical problems. Here are some of the research problems, in addition to those concerned with the core status characteristic theory, that have been studied in the standardized situation:

- Research on decision-making authority: This research investigates how individuals who possess different levels of decision-making authority (decision-maker vs. advisor) and different expectation states are influenced by the behavior of the other (Camilleri & Berger, 1967; Camilleri, Berger, & Conner, 1972).
- Research on the effects of sources of evaluations: Sources of evaluation are individuals with the right to evaluate the performances of an actor and are individuals whose evaluations also matter to the actor. This research has investigated how consistent and conflicting sources of evaluation can create different types of expectations, which in turn leads to different types of interactive behaviors (Webster & Sobieszek, 1974).

---

9. Note that in Rosenholtz’s intervention (as has been true of other such interventions), both the definition of the task situation and the structure of the task involve multi-abilities. In an important study, Goar and Sell (2005) showed that differentiation in the group’s interaction hierarchy can be reduced by multi-ability definition in the task situation while holding the structure of the task constant. See also Fişek (1991) for a theoretical formulation on complex task situations.

10. For further discussions of these and related issues on the external validity of experiments, see Lucas (2003b) and Webster (2005).

- Research on reward expectations: Reward expectations are expectations for the possession of rewards allocated in a status situation. Of particular interest are studies on what has been defined as the “reverse effect” in the status value theory of distributive justice (Berger et al., 1972). This is a theoretical argument that under appropriate conditions the allocation of rewards, in and of itself, creates task expectations (see Cook, 1975; Harrod, 1980; Stewart & Moore, 1992).
- Research on the transfer of status interventions: Status interventions are techniques that enable an individual to overcome invidious effects of status categorizations. This research has investigated the transfer of such interventions from an original status occupant to a second, from an original task to a new task, and from an original status category to a second and different category (e.g., race to educational attainment) (Berger, Fisek, & Norman, 1989; Lockheed & Hall, 1976; Markovsky, Smith, & Berger, 1984; Prescott, 1986; Pugh & Wahrman, 1983; Rashotte, 2006).
- Research on referent actors: This is research on how information about referent actors—individuals who are not interactants—affects the behavior of interactants in an immediate situation of action. An example is the effect on the behavior of a female (or male) actor in a mixed gender group with information that females (or males) in the past have outperformed members of the opposite gender on the same task that now confronts them (Freese, 1974; Wagner et al., 1986).
- Research on second-order expectations: If we consider the expectations that a first actor holds for self and a second actor as *first-order expectations*, then the expectations that the first actor believes the second holds for self and the first actor are labeled as *second-order expectations*. This research investigates the conditions under which second-order expectations are transformed into or affect the structure of first-order expectations (Fisek, Berger, & Moore, 2002; Kalkhoff, Younts, & Troyer, 2011; Moore, 1985; Troyer & Younts, 1997; Webster et al., 2004; Whitmeyer, Webster, & Rashotte, 2005).
- Research on the effects of sentiments on status behavior: This research is concerned with the conditions and processes by which positive and negative sentiments (e.g., patterns of likes and dislikes) either accentuate or attenuate the differentiation generated by status inequalities (Bianchi, 2005; Driskell & Webster, 1997; Fisek & Berger, 1998; Lovaglia & Houser, 1996; Shelly, 1993).
- Research on status cues: Status cues are behaviors, which may be expressive and nonverbal, that communicate information on the high or low performance capacities of an actor (e.g., rates of fluency and nonfluencies of speech). This research is concerned with the processes by which status cues affect the formation of interaction hierarchies in groups and, in turn, how status expectation differences affect the display of different types of status cues (Foddy & Riches, 2000; Rainwater, 1987; Riches & Foddy, 1989; Tuzlak & Moore, 1984; Webster & Rashotte, 2006).

- Research on status legitimization: A legitimated status order is one in which: (1) generalized deferential relations are prescriptive; (2) these relations have become embedded in power and prestige behaviors (e.g., the high-status individual should control the group's time and attention); and (3) there is the presumption that there will be collective support in maintaining the existing status order. This research is concerned with the emergence of legitimated status hierarchies and its consequences on group behavior ([Berger, Ridgeway, Fisek, & Norman, 1998](#); [Kalkhoff, 2005](#); [Ridgeway & Berger, 1986](#)).
- Research on multiple standards: This research is concerned with the use of different standards to evaluate the same performances of individuals who differ in status positions. The activation of multiple standards is viewed as a mechanism that in general operates to maintain existing status distinctions ([Foschi, 1996, 2000](#)).
- Research on social identity and status characteristic processes: This research is concerned with the effect of social identity and status characteristic processes examined singly as well as the combined effect of these processes on behavior as studied in the same experimental situation ([Kalkhoff & Barnum, 2000](#)).

## E SES and Other Research Sites

We never entertained the idea that the SES would be the only source of data relevant to theories in the expectation states program, and in fact it has not been the only source. Data relevant to theories in the program have come from a wide variety of research settings, including the following:

- Open-interaction studies ([Gallagher, Gregory, Bianchi, Hartung, & Harkness, 2005](#); [Propp, 1995](#); [Shelly & Munroe, 1999](#); [Skvoretz, Webster, & Whitmeyer, 1999](#)): Usually, these are studies of problem-solving groups whose members confront each other in face-to-face interaction.
- Studies in simulated organizational settings ([Hysom & Johnson, 2006](#); [Johnson, 2003](#); [Lucas, 2003a](#)).
- Actual on-site studies—for example, in classroom settings ([Cohen & Lotan, 1997b](#)), research and development teams ([Cohen & Zhou, 1991](#)), and on New York City Police teams ([Gerber, 2001](#)).
- Studies in hypothetical task situations ([Balkwell, Berger, Webster, Nelson-Kilger, & Cashen, 1992](#); [Fisek & Hysom, 2004](#); [Shelly, 2001](#)): Typically, these are studies in which the subject is asked to “imagine himself” in a social situation and then describe how he would react. This contrasts with studies in SES, in which the subject finds himself in a social situation created by the researcher, who also then observes and assesses the subject’s reactions.

- Studies that involve social applications and social inventions: Such studies cover a very broad range of settings from Israeli soccer teams ([Yuchtman-Yaar & Semyonov, 1979](#)) to breast cancer nursing wards ([Ludwick, 1992](#)).
- Studies in other kinds of experimental situations ([Dovidio, Brown, Heltmann, Ellyson, & Keating, 1988](#); [Driskell, Olmstead, & Salas, 1993](#)).
- Studies involving confederates ([Ridgeway & Erickson, 2000](#)): These are studies in which confederates are used to induce subordinate or superordinate behavior patterns on the part of the subjects.

Thus, for example, there is a status legitimization study done within SES ([Kalkhoff, 2005](#)) and another status legitimization study in a markedly different highly controlled experimental situation ([Johnson, 2003](#)). There is research on the “reverse process” (reward allocations leading to task expectations) done using the SES ([Cook, 1975](#)) and not using the SES ([Bierhoff, Buck, & Klein, 1986](#)) that yield comparable findings. In addition, there are research results from status cues studies in the SES ([Riches & Foddy, 1989](#)) as well as from the Nemeth–Wachtler jury-judging situation ([Mohr, 1986](#)) that are interrelated by theoretical research on the status cues formulation ([Fisek, Berger, & Norman, 2005](#); [Berger, Webster, Ridgeway, & Rosenholtz, 1986](#)).

One of the most interesting examples of research that involves a site other than SES, which directly tests an expectation state theory, is a study by [Dovidio and colleagues \(1988\)](#). This study was done in a setting they developed that involves two-person open-interaction groups whose members are working on short problem-solving tasks.

A prediction from the status characteristics theory (a major branch of the expectation states program) is that if a male and a female are working on a task that is not initially relevant to gender (a neutral task), then the male will have an expectation advantage over the female. If they are working on a masculine-typed task, the male’s expectation advantage will be even greater than when they are working on a neutral task; if they are working on a feminine-typed task, the expectation advantages will be reversed so as to favor the female. Furthermore, both observable power and prestige and status cues behaviors are predicted to be functions of one actor’s expectation advantage over a second actor.

We had actually considered doing research in SES to test for these arguments, but to our surprise and pleasure, we found that it already had been done by Dovidio and colleagues, and that they had obtained results that were exactly what we had predicted. They found that when working on gender-neutral tasks, males initiated more speech, spoke more, made more eye contact when speaking, and gestured more than females. These inequalities favoring males increased on masculine-typed tasks and, as predicted, were reversed when groups were working on feminine-typed tasks.

The fact that other research settings and methods have been used to test and apply the theories in the expectation program clearly adds to the credibility of these theories ([Cohen, 2003](#)).

## V CONCLUDING COMMENTS

Although the major components of the standardized situation had been developed by the early 1960s, graduate students and I (including Thomas Conner, Hamit Fisek, Murray Webster, James Moore, Barbara Meeker, Gordon Lewis, Karen Cook, and Ruth Cronkite) continued to work on refining the situation throughout the 1960s and early 1970s. We continued to make changes in the basic scenario to strengthen task and collective orientation by introducing the notion of a critical choice situation, and we shifted from using unit feedback on the individual's performances in manipulating expectations to simply handing out final scores to subjects. In fact, I began to wonder how little information we could give subjects and still manipulate their expectations, but we never pursued that problem. Eventually, a laboratory manual for doing research in SES was prepared by [Cook, Cronkite, and Wagner \(1974\)](#) and was subsequently revised by [Wagner and Harris \(1993\)](#).

By the end of the 1960s, B. P. Cohen introduced the use of video techniques in the SES, which gave us additional resources in our research ([Cohen, Kiker, & Kruse, 1969](#)). Among other things, by enabling us to present major portions of experimental procedures and information on videotape, it allows us to reduce the variability of theoretically irrelevant features in our experiments (e.g., variations in the behavior of the host experimenter). Subsequently, [Foschi \(1996\)](#) introduced a computerized version of the SES, and still later Lisa [Troyer \(1999, 2000\)](#) developed a second and [Webster et al. \(2004\)](#) developed a third computerized version of the situation. We recognized at the time that these innovations introduced major changes in the basic standardized situation. Research by [Troyer \(2001, 2002\)](#) and by [Kalkhoff and Thye \(Kalkhoff & Thye, 2006; Thye & Kalkhoff, 2009\)](#) has been concerned with examining the properties of the situation and, in particular, the variations in behavior introduced by the video and computerized versions.<sup>11</sup>

I think it is fair to say that we have witnessed major advances in our theoretical and empirical knowledge of status processes since the early 1960s (see [Berger & Webster, 2006; Wagner & Berger, 2002](#)). In these advances, the SES has played an important (but certainly not an exclusive) role. It has provided us with a sizeable quantity of comparable and relatively precise information that has facilitated the growth and testing of theories within the expectation states program. However, we should also observe that status research during this period has provided us with information such as that on status cues behaviors ([Fisek et al., 2005](#)), that on status

---

11. On the basis of a meta-analysis of 26 expectation states experiments, [Kalkhoff and Thye \(2006\)](#) documented the fact that the two situational parameters of the graph theory vary as between experiments conducted in the basic standardized situation (the standardized situation whose history is described in this paper), the video version of the SES, and the Foschi computerized version of the situation. They offer explanations to account for these effects, as well as effects due to the number of experimental trials and sample size. The analysis provides systematic information that researchers can use in planning future expectation states studies in standardized experimental settings.

latency effects (Conner, 1977, 1985), or that on status relations to acoustic behaviors (Gallagher et al., 2005). This information can be used to develop more elaborate as well as completely new versions of the standardized experimental situation for future research on expectation states and related social processes.

## APPENDIX

### The Structure of Status Characteristics Theory

#### I. Metatheoretical Components

Working strategies that are involved in defining theoretical problems and in formulating concepts and principles for their solution—for example, the concept of a status-organizing process in the status characteristics theory (Figure 12.1).

#### II. Theoretical Components

- A. *Scope conditions*: Statements that describe in abstract terms the social conditions under which the theory is assumed to hold.
- B. *Concepts and principles*: A set of abstract concepts and general principles that describe the operation (activation, evolution, and deactivation) of the social processes with which the theory is concerned. For status characteristics theory, see Figure 12.1.
- C. *Logical and mathematical structure*: A logical and mathematical structure within which the concepts and general principles are formulated in an interrelated and consistent structure (e.g., graph formulation in status characteristics theory).
- D. *General and specific derivations*: Statements about the relevant social processes that derive from the concepts and principles of the theory using the embedded logical or mathematical structure. General derivations describe behaviors in classes in situations (e.g., effects of increasing the number, relevance, or consistency of status distinctions on status behaviors). Specific derivations that describe behaviors for specific status situations.

#### III. Theory-Based Empirical Models

- A. *Instantiation of the theory*: This involves factual information or assumptions that relate abstract concepts in the theory to concrete conditions and phenomena—for example, information or assumptions to the effect that race or gender for a given population at a given time is a diffuse status characteristic.
- B. *Specification of the theory*: This involves information on the specific initial conditions that hold true of a given situation for which predictions are to be made (number of actors involved, type and states of status characteristics they possess, etc.).
- C. *Observations and the theory*: This involves assumptions and information that relate abstract theory to empirical observations. These include

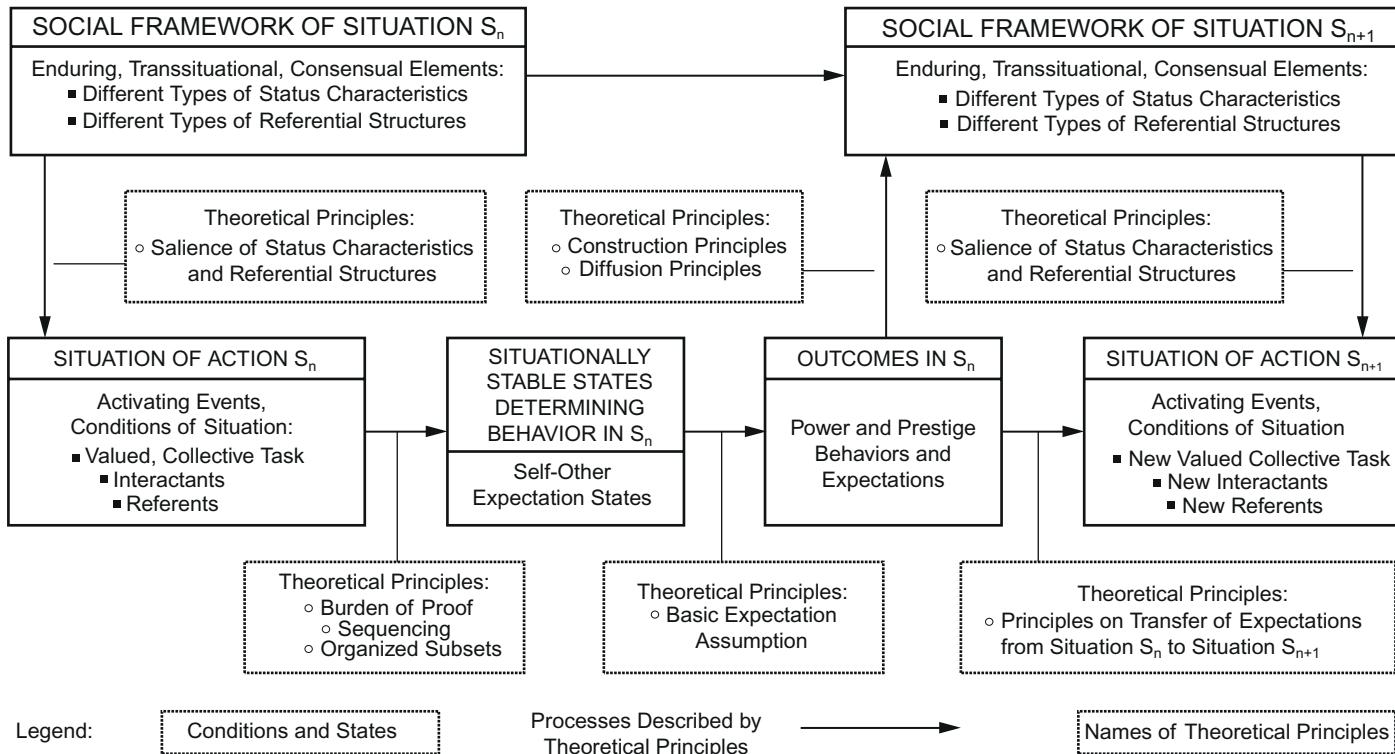


FIGURE 12.1 Status characteristics theory as a state organizing process. Legend: a=n. From Wagner and Berger (2002).

*coordinating assumptions*, as in those relating the abstract concept of “influence” in status theory with change in behavior in the standardized experimental situation. They also include *translation assumptions*, as in those that relate “expectation advantage” (theoretical) to the “proportion of stay responses” (observational). The latter assumption involves situational parameters whose values must be estimated to make specific behavioral predictions.

- D. *Simplifying assumptions*: One or more assumptions used to simplify representation of status situations or to simplify analysis and computations using status theory (e.g., that certain status task connections beyond a specified length are so weak that they can be ignored).

## ACKNOWLEDGMENTS

I acknowledge support from the Hoover Institution for work on this chapter. I express my appreciation to Robert Shelly, Jane Sell, Murray Webster, and Morris Zelditch for their comments and suggestions on the earlier version of this chapter.

## REFERENCES

- Bales, R. F. (1953). The equilibrium problem in small groups. In T. Parsons, R. Bales, & E. H. Shils (Eds.), *Working papers in the theory of action* (pp. 111–161). Glencoe, IL: Free Press.
- Bales, R. F., & Slater, P. (1955). Role differentiation in small decision-making groups. In T. Parsons, & R. F. Bales (Eds.), *Family, socialization and interaction process*. Glencoe, IL: Free Press.
- Bales, R. F., Stodbecht, F. L., Mills, T. M., & Roseborough, M. E. (1951). Channels of communication in small groups. *American Sociological Review*, 16, 461–468.
- Balkwell, J. W. (1991). Status characteristics and social interaction: An assessment of theoretical variants. *Advances in Group Processes*, 8, 135–176.
- Balkwell, J. W., Berger, J., Webster, M., Jr., Nelson-Kilger, M., & Cashen, J. (1992). Processing status information: Some test of competing theoretical arguments. *Advances in Group Processes*, 9, 1–20.
- Bavelas, A. (1950). Communication patterns in task-oriented groups. *Journal of the Acoustical Society of America*, 22, 725–730.
- Berger, J. (1958). *Relations between performance, rewards, and action-opportunities in small groups*. Ph.D. dissertation, Cambridge, MA: Harvard University.
- Berger, J. (1960). *An investigation of processes of role-specialization in small problem-solving groups*. Proposal funded by The National Science Foundation (July).
- Berger, J. (1974). Expectation states theory: A theoretical research program. In J. Berger, T. L. Conner, & M. H. Fisek (Eds.), *Expectation status theory: A theoretical research program* (pp. 3–22). Cambridge, MA: Winthrop.
- Berger, J., Cohen, B. P., & Zelditch, M., Jr. (1966). Status characteristics and expectation states. In J. Berger, M. Zelditch Jr., & B. Anderson (Eds.), *Sociological theories in progress* (pp. 29–46). Boston: Houghton Mifflin.
- Berger, J., Cohen, B. P., & Zelditch, M., Jr. (1972). Status characteristics and social interaction. *American Sociological Review*, 37, 241–255.

- Berger, J., & Conner, T. L. (1966). *Performance expectations and behavior in small groups* (Technical Report No. 18) Stanford, CA: Laboratory for Social Research, Stanford University.
- Berger, J., & Conner, T. L. (1969). Performance expectations and behavior in small groups. *Acta Sociologica*, 12, 186–197.
- Berger, J., & Fisek, M. H. (1970). Consistent and inconsistent status characteristics and the determination of power and prestige orders. *Sociometry*, 33, 278–304.
- Berger, J., & Fisek, M. H. (1974). A generalization of the theory of status characteristics and expectation states. In J. Berger, T. L. Conner, & M. H. Fisek (Eds.), *Expectation states theory: A theoretical research program* (pp. 163–205). Cambridge, MA: Winthrop.
- Berger, J., Fisek, M. H., & Crosbie, P. V. (1970). *Multi-characteristics status situations and the determinations of power and prestige orders*. (Technical Report No. 35) Stanford, CA: Laboratory for Social Research, Stanford University.
- Berger, J., Fisek, M. H., & Norman, R. Z. (1989). The evolution of status expectations: A theoretical extension. In J. Berger, M. Zelditch Jr., & B. Anderson (Eds.), *Sociological theories in progress: New formulations* (pp. 100–130). Newbury Park, CA: Sage.
- Berger, J., Fisek, M. H., Norman, R. Z., Zelditch, M., Jr. (1977). *Status characteristics and social interaction: An expectation states approach*. New York: Elsevier.
- Berger, J., Norman, R. Z., Balkwell, J. W., & Smith, R. F. (1992). Status inconsistency in task situations: A test of four status processing principles. *American Sociological Review*, 57, 843–855.
- Berger, J., Ridgeway, C. L., Fisek, M. H., & Norman, R. Z. (1998). The legitimization and delegitimation of power and prestige orders. *American Sociological Review*, 63, 379–405.
- Berger, J., & Snell, J. L. (1961). *A stochastic theory for self-other expectations* (Technical Report No. 1) Stanford, CA: Laboratory for Social Research, Stanford University.
- Berger, J., & Webster, M., Jr. (2006). Expectations, status, and behavior. In P. J. Burke (Ed.), *Contemporary social psychological theories* (pp. 268–300). Stanford, CA: Stanford University Press.
- Berger, J., Webster, M., Jr., Ridgeway, C. L., & Rosenholtz, S. (1986). Status cues, expectations, and behavior. *Advances in Group Processes*, 3, 1–22.
- Berger, J., & Zelditch, M., Jr. (Eds.) (1993). *Theoretical research programs: Studies in the growth of theory*. Stanford, CA: Stanford University Press.
- Bianchi, A. (2005). Rejecting others' influence: Negative sentiment and status in task groups. *Sociological Perspectives*, 47(4), 339–355.
- Bierhoff, H. W., Buck, E., & Klein, R. (1986). Social context and perceived justice. In H. W. Bierhoff, R. L. Cohen, & J. Greenberg (Eds.), *Justice in social relations* (pp. 165–185). New York: Plenum.
- Camilleri, S. F., & Berger, J. (1967, December). Decision-making and social influence: A model and an experimental test. *Sociometry*, 30, 367–378.
- Camilleri, S. F., Berger, J., & Conner, T. L. (1972). A formal theory of decisionmaking. In J. Berger, M. Zelditch Jr., & B. Anderson (Eds.), *In Sociological theories in progress; Vol. 2*. Boston: Houghton Mifflin.
- Cohen, B. P. (2003). Creating, testing, and applying social psychological theories. *Social Psychology Quarterly*, 66(1), 5–16.
- Cohen, B. P., Kiker, J. E., & Kruse, R. J. (1969). *The use of closed circuit television in expectation experiments* (Technical Report No. 29). Stanford, CA: Laboratory for Social Research, Stanford University.
- Cohen, E. G. (1982). Expectation states and interracial interaction in school settings. *Annual Review of Sociology*, 8, 209–235.
- Cohen, E. G., & Lotan, R. A. 1995a. Producing equal-status interaction in the heterogeneous classroom. *American Educational Research Journal*, 32, 99–120.
- Cohen, E. G., & Lotan, R. A. 1997a. *Working for equity in heterogeneous classrooms*. New York: Columbia University Teachers College Press.

- Cohen, B. P., & Zhou, X. (1991). Status processes in enduring work groups. *American Sociological Review*, 56, 179–188.
- Cohen, E. G., & Lotan, R. A. (1995a). Producing equal-status interaction in the heterogeneous classroom. *American Educational Research Journal*, 32, 99–120.
- Cohen, E. G., & Lotan, R. A. (1997a). *Working for equity in heterogeneous classrooms*. New York: Columbia University Teachers College Press.
- Cole, S. (2001). Why sociology doesn't make progress like the natural sciences. In S. Cole (Ed.), *What's wrong with sociology?* (pp. 37–60). New Brunswick, NJ: Transaction Books.
- Conner, T. L. (1964). *Three tasks for use in laboratory small-group experiments* (Technical Report). Stanford, CA: Laboratory for Social Research, Stanford University.
- Conner, T. L. (1965). *Continual disagreement and the assignment of self-other performance expectations*. Unpublished Ph.D. dissertation, Department of Sociology, Stanford University.
- Conner, T. L. (1977). Performance expectations and the initiation of problem solving attempts. *Journal of Mathematical Sociology*, 1977, 187–198.
- Conner, T. L. (1985). Response latencies, performance expectations, and interaction patterns. In J. Berger, & M. Zelditch Jr., (Eds.), *Status, rewards, and influence: How expectations organize behavior* (pp. 189–214). San Francisco: Jossey-Bass.
- Cook, K. S. (1975). Expectations, evaluations, and equity. *American Sociological Review*, 40, 372–388.
- Cook, K. S., Cronkite, R., & Wagner, D. G. (1974). *Laboratory for social research manual for experiments in expectation state theory*. Stanford, CA: Laboratory for Social Research, Stanford University.
- Dovidio, J. F., Brown, C. E., Heltmann, K., Ellyson, S. L., & Keating, C. F. (1988). Power displays between women and men in discussions of gender-linked tasks: A multichannel study. *Journal of Personality and Social Psychology*, 55, 580–587.
- Driskell, J. E., Olmstead, B., & Salas, E. (1993). Task cues, dominance cues, and influence in task groups. *Journal of Applied Psychology*, 78, 51–60.
- Driskell, J. E., & Webster, M., Jr. (1997). Status and sentiment in task groups. In J. Szmatka, J. Skvoretz, & J. Berger (Eds.), *Status, network, and organization* (pp. 179–200). Stanford, CA: Stanford University Press.
- Fisek, M. H. (1991). Complex task structures and power and prestige orders. *Advances in Group Processes*, 8, 115–134.
- Fisek, M. H., & Berger, J., (1998). Sentiment and task performance expectations. *Advances in Group Processes*, 15, 23–40.
- Fisek, M. H., Berger, J., & Moore, J. C., Jr. (2002). Evaluations, enactment and expectations. *Social Psychology Quarterly*, 65, 329–345.
- Fisek, M. H., Berger, J., & Norman, R. Z. (2005). Status cues and the formation of expectations. *Social Science Research*, 34, 80–102.
- Fisek, M. H., & Hysom, S. J. (2004). Status characteristics and reward expectations: Test of a model. In *Paper presented at the annual meeting of the American Sociological Association, San Francisco, August 14–17*.
- Fisek, M. H., Norman, R. Z., & Nelson-Kilger, M. (1992). Status characteristics and expectation states theory: *A priori* model parameters and test. *Journal of Mathematical Sociology*, 16, 285–303.
- Foddy, M., & Riches, P. (2000). The impact of task and categorical cues on social influence: Fluency and ethnic accent as cues to competence in task groups. *Advances in Group Processes*, 17, 103–130.
- Foschi, M. (1996). Double standards in the evaluation of men and women. *Social Psychology Quarterly*, 59, 237–254.
- Foschi, M. (2000). Double standards for competence: Theory and research. *Annual Review of Sociology*, 26, 21–42.
- Freese, L. (1974). Conditions for status equality. *Sociometry*, 37, 147–188.

- Gallagher, T. J., Gregory, S. W., Jr., Bianchi, A. J., Hartung, P. J., & Harkness, S. (2005). Examining medical interview asymmetry using the expectation states approach. *Social Psychology Quarterly*, 68, 187–203.
- Gerber, G. L. (2001). *Women and men police officers: Status, gender, and personality*. Westport, CT: Praeger.
- Goar, C., & Sell, J. (2005). Using task definition to modify racial inequality within task groups. *Sociological Quarterly*, 46, 525–543.
- Harrod, W. (1980). Expectations from unequal rewards. *Social Psychology Quarterly*, 43, 126–130.
- Harvey, O. J. (1953). An experimental approach to the study of status relations in informal groups. *American Sociological Review*, 18, 357–367.
- Heinecke, C., & Bales, R. F. (1953). Developmental trends in small groups. *Sociometry*, 16, 7–38.
- Hysom, S. J., & Johnson, C. (2006). Leadership structures in same-sex task groups. *Sociological Perspectives*, 49, 391–410.
- Johnson, C. (2003). Consideration of legitimacy processes in teasing out two puzzles in the status literature. In S. R. Thye, & J. Skvoretz (Eds.), *Advances in group processes, power and status* (pp. 251–284). Greenwich, CT: JAI Press.
- Kalkhoff, W. (2005). Collective validation in multi-actor task groups: The effects of status differentiation. *Social Psychology Quarterly*, 68, 57–88.
- Kalkhoff, W., & Barnum, C. (2000). The effects of status-organizing and social identity processes on patterns of social influence. *Social Psychology Quarterly*, 63, 95–115.
- Kalkhoff, W., & Thye, S. (2006). Expectation states theory and research: New observations from meta-analysis. *Sociological Research and Methods*, 35, 219–249.
- Kalkhoff, W., Younts, C. W., & Troyer, L. (2011). Do second-order expectations transfer to new groups and tasks? An expectation states approach. *Social Psychology Quarterly*, 74, 267–290.
- Leavitt, H. J. (1950). Some effects of certain communication patterns on group performance. *Journal of Abnormal and Social Psychology*, 46, 38–50.
- Lockheed, M. E., & Hall, K. P. (1976). Conceptualizing sex as a status characteristic: Applications to leadership training strategies. *Journal of Social Issues*, 32, 111–124.
- Lovaglia, M. J., & Houser, J. A. (1996). Emotional reactions and status in groups. *American Sociological Review*, 61, 867–883.
- Lucas, J. W. (2003a). Status processes and the institutionalization of women as leaders. *American Sociological Review*, 68, 464–480.
- Lucas, J. W. (2003b). Theory-testing, generalization, and the problem of external validity. *Sociological Theory*, 21, 236–253.
- Ludwick, R. (1992). Registered nurses' knowledge and practices of teaching and performing breast exams among elderly women. *Cancer Nursing*, 15, 61–67.
- Markovsky, B., Smith, R. F., & Berger, J. (1984). Do status interventions persist? *American Sociological Review*, 49, 373–382.
- Mohr, P. B. (1986). Demeanor, status cue or performance? *Social Psychology Quarterly*, 49, 228–236.
- Moore, J. C., Jr. (1965). *Development of the spatial judgment experimental task* (Technical Report No. 15). Stanford, CA: Laboratory for Social Research, Stanford University.
- Moore, J. C., Jr. (1985). Role enactment and self-identity. In J. Berger, & M. Zelditch Jr., (Eds.), *Status, rewards, and influence: How expectations organize behavior* (pp. 262–316). San Francisco: Jossey-Bass.
- Norman, R. Z., Smith, R. F., & Berger, J. (1988). The processing of inconsistent status information. In M. Webster Jr., & M. Foschi (Eds.), *Status generalization: New theory and research* (pp. 169–187). Stanford, CA: Stanford University Press.
- Platt, J. R. (1964). Strong inference. *Science*, 146, 347–353.

- Prescott, W. S. (1986). *Expectation states theory: When do interventions persist?* Unpublished manuscript, Hanover, NH: Dartmouth College.
- Propp, K. M. (1995). An experimental examination of biological sex as a status cue in decision-making groups and its influence on information use. *Small Group Research*, 26, 451–474.
- Pugh, M. D., & Wahrman, R. (1983). Neutralizing sexism in mixed-sex groups: Do women have to be better than men? *American Journal of Sociology*, 88, 736–762.
- Rainwater, J. A. (1987). *Status cues: A test of an extension of status characteristics theory*. Ph.D. dissertation, Stanford, CA: Department of Sociology, Stanford University.
- Rashotte, L. S. (2006). Controlling the status effects of gender. In *Paper presented at the annual meeting of the International Society of Political Psychology, Barcelona*.
- Riches, P., & Foddy, M. (1989). Ethnic accent as status cue. *Social Psychology Quarterly*, 52, 197–206.
- Ridgeway, C. L., & Berger, J. (1986). Expectations, legitimization, and dominance behavior in task groups. *American Sociological Review*, 51, 603–617.
- Ridgeway, C. L., & Erickson, K. G. (2000). Creating and spreading status beliefs. *American Journal of Sociology*, 106, 579–615.
- Rosenholtz, S. J. (1977). *The multiple ability curriculum: An intervention against the self-fulfilling prophecy*. Unpublished doctoral dissertation, Stanford, CA: Department of Sociology, Stanford University.
- Shelly, R. K. (1993). How sentiments organize interaction. *Advances in Group Processes*, 10, 113–132.
- Shelly, R. K. (2001). How performance expectations arise from sentiments. *Social Psychology Quarterly*, 64, 72–87.
- Shelly, R., & Munroe, P. (1999). Do women engage in less task behavior than men? *Sociological Perspectives*, 42, 49–67.
- Sherif, M., White, B. J., & Harvey, O. J. (1955). Status in experimentally produced groups. *American Journal of Sociology*, 60, 370–379.
- Skvoretz, J., Webster, M., Jr., & Whitmeyer, J. (1999). Status orders in task discussion groups. *Advances in Group Processes*, 16, 199–218.
- Stewart, P., & Moore, J. C. (1992). Wage disparities and performance expectations. *Social Psychology Quarterly*, 55, 78–85.
- Thye, S., & Kalkhoff, W. (2009). Seeing the forest through the trees: An updated meta-analysis of expectation states research. *Current Research in Social Psychology*, 15(1), 1–14.
- Troyer, L. (1999). MacSES v. 5.0. Unpublished software manual.
- Troyer, L. (2000). MacSES v. 5.0. Unpublished software manual.
- Troyer, L. (2001). Effects of protocol differences on the study of status and social influence. *Current Research in Social Psychology*. Available online at <http://www.uiowa.edu/~grpproc>.
- Troyer, L. (2002). The relation between experimental standardization and theoretical development in group processes research. In J. Szmakta, M. Lovaglia, & K. Wysierska (Eds.), *The growth of social knowledge: Theory, simulation, and empirical research in group processes* (pp. 131–147). Westport, CT: Praeger.
- Troyer, L., & Younts, C. W. (1997). Whose expectations matter? The relative power of first-order and second-order expectations in determining social influence. *American Journal of Sociology*, 103, 692–732.
- Tuzlak, A., & Moore, J. C., Jr. (1984). Status, demeanor, and influence: An empirical assessment. *Social Psychology Quarterly*, 47, 178–183.
- Wagner, D. G., & Berger, J. (1985). Do sociological theories grow? *American Journal of Sociology*, 90, 697–728.

- Wagner, D. G., & Berger, J. (2002). Expectation states theory: An evolving research program. In J. Berger, & M. Zelditch Jr. (Eds.), *New directions in contemporary sociological theory* (pp. 41–76). Lanham, MD: Rowman & Littlefield.
- Wagner, D. G., Ford, R. S., & Ford, T. W. (1986). Can gender inequalities be reduced? *American Sociological Review*, 51, 47–61.
- Wagner, D. G., & Harris, R. O. (1993). *Manual for experimenters in expectation states theory* (3rd ed., technical report). Albany, NY: Group Processes Research Office, Department of Sociology, University at Albany.
- Walker, H. A., & Cohen, B. P. (1985). Scope statements: Imperatives for evaluating theory. *American Sociological Review*, 40, 288–301.
- Webster, M., Jr. (1969). Sources of evaluations and expectations for performance. *Sociometry*, 32, 243–258.
- Webster, M., Jr. (2005). Laboratory experiments in social science. *Encyclopedia of Social Measurement*, 2, 423–433.
- Webster, M., Jr., & Entwistle, D. R. (1974). Raising children's expectations for their own performance: A classroom application. In J. Berger, T. L. Conner, & M. H. Fisek (Eds.), *Expectation states theory: A theoretical research program* (pp. 211–243). Cambridge, MA: Winthrop.
- Webster, M., Jr., & Rashotte, L. S. (2006). How behavior affects performance expectations. In *Paper presented at the annual meeting of the International Society of Political Psychology, Barcelona*.
- Webster, M., Jr., & Sobieszek, B. I. (1974). *Sources of self-evaluation: A formal theory of significant others and social influence*. New York: Wiley.
- Webster, M., Jr., Whitmeyer, J. M., & Rashotte, L. S. (2004). Status claims, performance expectations, and inequality in groups. *Social Science Research*, 33, 724–745.
- Whitmeyer, J. M., Webster, M., Jr., & Rashotte, L. S. (2005). When status equals make status claims. *Social Psychology Quarterly*, 68, 179–186.
- Whyte, W. F. (1943). *Street corner society*. Chicago: Chicago University Press.
- Yuchtman-Yaar, E., & Semyonov, M. (1979). Ethnic inequality in Israeli schools and sports: An expectation-states approach. *American Journal of Sociology*, 85, 576–590.
- Zelditch, M., Jr., Lauderdale, P., & Stublarec, S. (1975). *How are inconsistencies in status and ability resolved?* (Technical Report No. 54). Stanford, CA: Laboratory for Social Research, Stanford University.
- Zelditch, M., Jr., Lauderdale, P., & Stublarec, S. (1980). How are inconsistencies between status and ability resolved? *Social Forces*, 58, 1025–1043.

## Chapter 13

# Experimental Political Science

Rose McDermott

*Brown University, Providence, Rhode Island*

## I INTRODUCTION

The Women's Health Initiative (WHI), designed to examine the causes of morbidity and mortality in postmenopausal women, was one of the most expensive publicly funded experiments in U.S. history, costing more than \$1 billion. Although the original study comprised four different elements, the studies involving the use of estrogen generated the most publicity and resulted in the WHI being terminated early due to concerns that those in the treatment group were at increased risk for cardiovascular disease. Millions of women stopped their hormone replacement treatment as a result. However, 10 years later, it has become clear, based on the results of subsequent studies and re-analysis of the original data, that the protective effects of estrogen, particularly for younger women, remain strong. Indeed, Lobo (2013) states:

*It has been argued that in the 10 years since WHI, many women have been denied HT, including those with severe symptoms, and that this has significantly disadvantaged a generation of women. Some reports have also suggested an increased rate of osteoporotic fractures since the WHI. (p. 1171)*

Therefore, in addition to affecting the health and comfort of millions of women, and costing taxpayers \$1 billion, these experiments affected health policy and public health funding priorities in significant ways to the detriment of public welfare. This confound painfully illustrates the reason why experiments provide an important albeit challenging method of social inquiry: sufficiently large populations can display precisely the kind of variance in behavior and outcome necessary to disentangle the impact of various causes on particular outcomes of interest, but they can also pose huge costs and risks for the public at large if such studies are poorly designed or improperly interpreted. Significantly, the senior authors of the study were medical professionals, not social scientists. In addition, many of the obvious demographic features that remain standard for social scientists, such as education and geography, were largely left out of the original analysis. Even more interesting, many of the original interpretations of

the results, and consequent health policy recommendations, ran clearly contrary to the data reported in the studies; it simply was the case that too few people paid attention to, or understood, how to properly interpret that data. This points to the critically important role that social scientists can and should play in large experimental studies, including medical ones, that involve public funds and public health policy.

Although the use of experimental methodology by political scientists has increased dramatically in approximately the past decade (Druckman, Green, Kuklinski, & Lupia, 2011), some scholars still do not embrace its viability for the nondiscursive and nonnative questions that preoccupy the discipline (Smith, 2002). However, Nobel Laureate Elinor Ostrom (2002), among others, called for a greater integration of experimentation, along with other methodologies, to more successfully address enduring questions in political science, such as collective action, voting, and multicultural group behavior. Although most experimental applications in political science have focused on American politics and voting behavior, increasingly scholars in international relations and comparative politics have become interested in incorporating experiments as one potentially useful method to employ in investigating questions and problems of interest, and the use of field experiments in particular has become increasingly common in these subfields.

Although topics related to voting continue to dominate the use of experiments in political science, and Americanists have embraced their usage most heartily, experiments still do not represent the most common method by which to examine any given topic in political science. In other words, despite the recent appearance of a new journal dedicated to their use—the *Journal of Experimental Political Science*—and although experiments continue to increase in number, legitimacy, sophistication, and breadth of application, they still have yet to become a dominant method in the field at large. Indeed, in the most recent survey of more than 1,000 international relations scholars conducted by the editors of *Security Studies* (Peterson & Tierney; 2005), less than 4% reported using experimental methodology at all in their work. In a survey of international relations in the U.S. academy, experimental methods still fell under the “other” category (Maliniak, Oakes, Peterson, & Tierney, 2011).

This chapter briefly examines some of the historical developments in experimental political science. Next, some more contemporary examples illustrate the ways in which innovative techniques can throw light on old problems in new ways. Finally, an assessment of the field, along with a discussion of future opportunities and challenges, is provided.

## II HISTORICAL DEVELOPMENT

The use of experiments in political science dates back at least to the work of Harold Gosnell in 1926. Starting in 1923, he conducted a field experiment to examine the reasons why people did not vote in the mayor’s election in Chicago.

In his first study, he reported on the effect of direct mail on voter turnout. Following up on an earlier study involving 6,000 personal interviews, Gosnell identified an additional 6,000 voters in Chicago residing in 12 districts. In each district, he randomly assigned half the voters to receive postcard reminders to register to vote, while the other half of citizens received no such intervention.

The first postcard, printed in several languages, resulted in 42% of 3,000 potential voters registering to vote, whereas only 33% of the 2,700 citizens who did not receive mail registered. A second follow-up postcard resulted in 56% of the 1,700 subjects being stimulated to register, whereas only 47% of the 1,770 unstimulated counterparts registered. This card had a note and a cartoon portraying nonvoters as slackers; the cartoon proved slightly more effective with women. Overall, 75% of those who received a card registered, whereas only 65% of those who did not receive mail registered. The impact of direct mail on actual votes varied across districts, and it was related to the strength of the local party organizations. This experiment provided a model for the development of future experimental work in political science for decades to come, in at least three ways.

First, it focused on issues relating to American voting behavior, still the most common topic for experimental investigation by political scientists (for a review, see [McDermott, 2002](#)). Second, it examined an applied issue of concern to academic as well as real-world policymakers and, sometimes, even the interested public. This approach continues today through work on topics such as voting; campaigns and elections; committee and jury decision-making; and problems relating to coordination, cooperation, bargaining, negotiation, and conflict resolution. Last, [Gosnell's \(1926\)](#) experiment examined a topic that had generated a great deal of interest among political scientists and had been studied widely from the perspective of other methodologies but without any causal consensus emerging. This tradition continues today, as political scientists often invoke experimental methodology to investigate topics that have received widespread attention while yielding either inconsistent or confusing results. The most influential modern trend in such voting research embeds experimental manipulations in large survey instruments ([Kuklinski et al., 1997](#)) to combine the control and internal validity offered by laboratory experiments with the generalization and external validity offered by nationally representative survey samples.

Starting relatively early in political science, many scholars believed that their experimental work was being unfairly rejected from established political science journals. In response to this concern, the journal *Experimental Study of Politics* was founded in 1971. However, the journal lasted only 4 years, at least partly because a great deal of the experimental work published in this title was not as sophisticated as work on similar topics being published in psychology journals. Indeed, many of the best experimental political scientists of the time published their work on voting in psychology journals or in edited volumes devoted to specific topics, such as media effects or race. After the demise of this journal, the number of experimental articles published in political science journals did increase over time but continued to be concentrated in

a group of specific journals, most notably those focused on American voting behavior, and *Political Psychology*. Predominant topics of interest continued to include voting, games such as prisoner's dilemma and game behavior, bargaining, committee rules and work, race, and media effects. Recently, the *Journal of Experimental Political Science* was launched, offering a new opportunity for scholars employing experimental methods to showcase their work; this journal explicitly endorses the publication of null results and seeks replication studies as well.

As the use of experimental methodology has broadened beyond voting in recent years, scholars in international relations and even comparative politics have begun to undertake such work as well. In international relations, such research has included studies involving simulated arms races, bargaining and negotiations (Deutsch, Epstein, Canavan, & Gumpert, 1967), war games and crisis simulations (Beer, Healy, Sinclair, & Bourne, 1987; Beer, Sinclair, Healy, & Bourne, 1995), and foreign policy analysis and decision-making (Geva, Mayhar, & Skorick, 2000; Mintz & Geva, 1993; Mintz, Geva, Redd, & Carnes, 1997). Increasingly, however, these scholars, in concert with those in comparative politics, are undertaking field experiments in various locations throughout the world. Many of these experiments involve the use of games adopted from behavioral experiments to examine various political attitudes and beliefs in cross-cultural contexts. Many of these experiments focus on topics such as foreign aid, foreign direct investment, humanitarian intervention, and security strategies following civil war.

Interestingly, some scholars in comparative politics have begun to undertake field experiments as well. Whereas some of these studies examine voting behavior in other countries (Wantchekon, 2003), others have adapted the methodology for investigating topics including cultural differences between ethnic groups in Africa (Posner, 2004) and nation building and public goods problems there as well (Miguel, 2004). Rick Wilson and Donna Bahry accomplished a remarkable undertaking in conducting a series of experimental laboratory economic games, including the ultimatum game, between and among several different ethnic groups in Russia at various locations (Bahry, Kosolopov, Kozyreva, & Wilson, 2005; Bahry & Wilson, 2006). These studies examined the nature of trust between ethnic groups and found, perhaps surprisingly, higher levels of inter-ethnic trust than previous observers appeared to expect.

Some of the most productive areas of experimental research have involved the study of public goods and free-rider problems drawn from economics to examine fundamentally political problems. John Orbell and colleagues (Dawes, Orbell, Simmons, & van de Kragt, 1986; van de Kragt, Orbell, & Dawes, 1983) have conducted a number of experiments from this perspective. In one set of small-group experiments, van de Kragt et al. examined the effect of communication on contributions to public goods. They found that communication resulted in an efficient production of the public good, whereas lack of communication resulted in failure to create the good more than one-third of the time.

In further work, Dawes et al. explored the effectiveness of various incentives on contributions to public goods. In three experiments, they discovered that enforcing contributions in a “fair share” agreement did increase contributions. [Ostrom, Walker, and Gardner \(1992\)](#) also investigated common-good problems experimentally as well to assess the effect of various strategies on developing credible commitments among contributors. These techniques, which included “covenants” allowing communication, “swords” offering opportunities for sanctions, and combinations of both, did result in self-governance without the presence of external enforcement.

### III RECENT EXAMPLES

Recent examples of the use of experiments in political science constitute three main types. First, experimental examination has accumulated and aggregated in certain areas, most notably voting behavior, but also in areas such as collective action. Second, experiments have recently provided some innovative ways to explore socially sensitive or previously inaccessible topics using new methods. The last involves the use of field experiments in an attempt to increase the ecological validity and generalization of experiments in international and comparative contexts. Each of these is discussed briefly in turn.

First, several topics in American voting behavior have been subjected to extended and repeated experimental tests and manipulations. Notable in this regard, and clearly inspired by the work originally undertaken by Gosnell, is the work of Donald Green and Alan Gerber on voter turnout. This work includes examinations of the effect of leafleting ([Gerber & Green, 2000a](#)), canvassing, direct mail and phone banks ([Gerber & Green, 2000b; Green, 2004](#)), and habit ([Gerber, Green, & Shachar, 2003](#)). Other scholars have expanded on the use of field experiments to examine other aspects of voting behavior, including race. In a 2002 study of Asian American voters in Los Angeles County, [Wong \(2005\)](#) found that both telephone calls and postcards increased voter turnout. In a similar but much larger study of Latino voters in 2002, [Ramirez \(2005\)](#) randomly assigned more than 465,000 Latino voters to receive direct mail, live phone calls, or robotic calls. In this study, contrary to the results Wong reported among Asians, Ramirez found that only live calls increased voter turnout among Latinos. More recent extensions have incorporated the use of cluster analysis into field experiments to study voting behavior ([Arceneaux, 2005](#)).

Turnout does not represent the only aspect of experimental investigation into voting. In fact, not all voting research takes place in the context of field experiments. Indeed, the majority of experimental work in voting takes place in laboratory settings. Such investigations include exploration into topics such as campaigns and elections, most notably the impact of television news ([Iyengar, Peters, & Kinder, 1982](#)) and negative advertising ([Anscombe, Iyengar, Simon, & Valentino, 1999; Anscombe, Iyengar, Simon, & Valentino, 1994](#)) on voter preference. Interesting work has also attempted to use experimental manipulations

to examine the way in which political advertising can cue and prime socially sensitive topics such as race to influence attitudes during campaigns ([Valentino, Hutchings, & White, 2002](#)).

Indeed, other areas of inquiry using experimental methods to explore various aspects of voting engage explicitly psychological models of human decision-making. Such work includes the burgeoning area of experimental investigations into the dynamics underlying framing effects ([Druckman, 2001a, 2001b](#)), as well as the affective forces that support political party identification ([Burden & Klofstad, 2005](#)).

A second way in which contemporary work in political science can be examined is through the lens of new technologies, which can be used to explore previously inaccessible or sensitive topics in new ways. Two research prospects deserve particular mention in this regard. First, Shanto Iyengar and colleagues have developed a very interesting and sophisticated design to manipulate similarity or familiarity or both. Using morphing technology, these authors integrate the face of the subject with that of noteworthy politicians or celebrities to examine the impact of familiarity and similarity on voter preference. In one study of similarity, [Bailenson, Garland, Iyengar, and Yee \(2006\)](#) found intriguing gender differences, such that men proved more likely than women to vote for a candidate who resembled them. Although a Democratic candidate provided the only stimulus in this particular study, political party identification did not appear to influence voter preference in this experiment.

Another version of this study examining familiarity used large, if somewhat biased, convenience samples derived from readership of the *Washington Post*. In fact, Iyengar agreed to provide one such online survey instrument for the newspaper per month for a year. In one study, more than 2,000 participants in both political parties demonstrated a decided preference for an unknown face morphed with that of Senator Hilary Clinton over the same face morphed with Senator John McCain. In a second study, which adapted the popular “whack-a-mole” game into “whack-a-pol,” Iyengar demonstrated partisan whacking. When subjects had to hit famous celebrities (Michael Jackson, Brad Pitt, and Angelina Jolie) or famous dictators (Joseph Stalin, Adolph Hitler, and Saddam Hussein), whacking appeared rapid and indiscriminate in nature. However, when subjects had to attack politicians (John Kerry and George Bush), people took much more time and care in whom they hit, demonstrating clear partisan bias in comparison to the other conditions, in which subjects were able to hit many more targets much faster because they did not care whom they whacked ([Iyengar, 2002](#)).

In another particularly telling demonstration, readers were asked how much federal assistance a victim of Hurricane Katrina should receive. In each case, the story remained the same, but the picture of the victim provided to readers varied by race. Those readers who saw a white victim advocated for more aid than those who received the black or Hispanic picture of the ostensible victim (for a review, see [Stanford Magazine, 2006](#)). Such innovative techniques offer unique and unprecedeted ways of examining socially sensitive topics such as

race in less obtrusive ways. Another example comes from work by Jennifer Eberhardt, a psychologist who demonstrated that subjects were more likely to give harsher sentences, including the death penalty, to perpetrators whose skin was experimentally manipulated to be darker in tone in the pictures presented to subjects (Eberhardt, Davies, Purdie-Vaughns, & Johnson, 2006).

A second area of innovative research involves the use of new technologies in brain mapping to investigate the previously inaccessible “black box” of decision-making. Three of these advances deserve particular mention. The first involves the use of functional magnetic resonance imaging (fMRI) technology to better understand the parts of the brain that are activated during particular judgments, feelings, or behaviors. Darren Schreiber and colleagues (Lieberman, Schreiber, & Ochsner, 2003) have used such techniques to explore the differences in political thinking and processing between experts and novices. The second uses electroencephalogram (EEG) technology to examine the timing of particular processes in the brain. Rick Wilson and colleagues (Wilson, Stevenson, & Potts, 2006) have used this strategy to explore the mental differences in subjects playing both dominant and mixed strategy games. Note that MRIs provide better information regarding the geography of brain activity, whereas EEGs offer more precise data on the timing of such events. A third uses other physiological measures such as galvanic skin response or eye blink to examine the speed and intensity of reactions to particular phenomena such as threat (Oxley et al., 2008). Recent research investigating the biological bases of human decision-making seeks to traction specific genetic and hormonal markers to examine their impact on various behaviors of interest, including aggression and cooperation.

Finally, the increasing prevalence of field experiments portends a powerful extension of the techniques and methods of laboratory experimentation, which privileges internal validity, to more generalized settings. Such an expansion offers the possibility of exploring topics of interest to scholars of international relations and comparative politics using experimental methodology. Although much of this work has yet to be published in journals, a great deal of energy is being invested in a variety of topics and locales. Interesting work in this area includes a natural experiment investigating electoral fraud in Armenia (Hyde, 2007), a field experiment assessing democracy promotion in Indonesia (Hyde, 2010), as well as work on corruption in Brazil (Weitz-Shapiro & Winters, 2010) and security sector reform following civil conflict in countries such as Liberia.

## IV ASSESSMENT AND CHALLENGES

There are several ways in which these new techniques might be effectively exploited to examine topics of interest and importance to political scientists across subfields. Methodological innovations alone cannot drive creativity, of course. Innovative technology should be harnessed in service of cohesive and comprehensive theoretical models of the problems and topics under investigation.

In addition, institutional and structural challenges, as well as incentive structures that often continue to privilege sole author work, confront any attempt to make the use of experiments more widespread in political science.

## V POTENTIALS FOR FUTURE WORK

There are at least three areas in which innovative technologies hold promise for creative experimental explorations in political science. First, the use of brain mapping technologies such as fMRI, EEG, and positron emission topography (PET) offer previously unprecedented access into the workings of the human mind. Although these technologies are relatively new, their increasingly widespread use in the cognitive neurosciences has led to an explosion of new discoveries in the operation and psychological dynamics underlying human thoughts, feelings, and behaviors.

For applications in the political realm, several possibilities in addition to the obvious ones relating to voter judgments of candidates appear particularly appealing. Because such technology can now be used in a synchronous manner, studies of strategic interaction seem an obvious first step for exploration. Such research might investigate structural and psychological factors that can affect bargaining, negotiation, reputation, status, power, and other topics related to achieving cooperative outcomes.

Three specific areas of application in this arena seem particularly promising. First, studies that examine the difference in process and outcome when monetary versus nonmonetary payoffs are under consideration could prove particularly illuminating for understanding political motivation in noneconomic realms. Essentially, all experimental research in behavioral economics uses monetary incentives to motivate subject participation and action. However, as psychologists, sociologists, and political scientists recognize, many people remain additionally, or even solely, motivated by nonmonetary considerations, especially social and status concerns and rewards. For example, Colin Camerer (personal communication) finds that subjects playing experimental games related to saving money for retirement learn at a different rate, and adopt different strategies, if they are playing for juice after having been made thirsty than when playing for money.

Comparing strategies when monetary versus social incentives are at play offers a particularly interesting way to investigate other forces that may decisively motivate human action in the political and social realms. [Sell, Griffith, and Wilson \(1993\)](#) found gender differences by resources as well. In two experiments, these authors examined the effect of the subject's gender, the gender composition of his or her group, and the resources involved within the context of social dilemma to explain gender differences in contributions. In this study, gender did not affect individual contributions when money was the relevant resource. However, when other social resources were involved, gender differences in fact emerged. This example

demonstrates how simple changes in the operationalization of various measures can either obfuscate or highlight the significance of various variables, including standard demographic ones.

Examining the differences in decision-making when people play against other people as opposed to playing against a computer may help provide a model for individual action and apathy in modeling such processes as citizen engagement, civic action and participation, civil disobedience, and processes surrounding the collective action problems inherent in instigating rebellion and revolution. It is doubtful individuals respond to bureaucracies in ways that resemble their response to computers. If individuals can personalize their experience of government, does it change the strategies and actions they take with regard to processes of civic engagement? Such questions might be illuminated through a comparison of such processes. The British government, for example, is trying to replace gross domestic product with a fundamentally more psychological notion of national well-being as a better indicator of health and development; examining how such measures might potentiate human capital offers one way to combine these last two areas of application into a united research agenda that can experimentally explore the interaction of technology, governance, and both economic and social development and equality.

Moreover, the impact of rhetoric and communication styles and strategies receives different treatment in various areas of academic and public political discourse. In American politics, such attention to “spin” appears ubiquitous, if often undertheorized. In international relations, the impact of such factors is often ignored, or it is taken for granted but assumed not to influence outcomes in any material fashion. But MRI technology might provide an avenue by which to examine the effect of communication on actions such as aggression and co-operation. It may be that certain types of individuals are more affected by such discourse than others. If such impact were predictable and discernible—based, for example, on genetic differences—it could influence strategies of public education in important arenas such as health policy, as well as have an impact on campaign strategies.

A second area of investigation that offers promise for advances in political science lies in the domain of genetic and biological influences on behavior. This arena encompasses at least three separate kinds of studies. The classic way to investigate such differences involves twin studies. Such examinations, along with other genetic studies, including genome-wide association studies, offer the standard way to determine the extent to which behavioral outcomes are relatively influenced by genetic and environmental factors. Twin studies might be employed to compare temperamentally influenced albeit socially complex characteristic traits of interest, such as hostility, trustworthiness, aggression, and competitiveness, to determine how much of each trait might be genetically informed or environmentally influenced. They may also be used to examine how different types of incentives or sanctions might affect different kinds of people in systematic or predictable ways.

Another way to investigate the genetic determinants of behavior involves DNA analysis, using cheek swabs or hair samples to divide populations into those who differ on some genetic variation of interest. For example, some recent discussion suggests that those with genetic mutations in arginine vasopressin (AVP) will exhibit different patterns of parenting, aggression, and fairness in economic allocation. Similarly, other important genetic markers can now be extracted from saliva, including monoamine oxidase (MAO) inhibitors. Understanding the etiology of, and consequences associated with, such differences may help illuminate important tendencies in behaviors such as aggression and impulsivity. Clearly, most complex social and political behaviors remain far too vast and interactive to be driven by simple mutations on single genes. However, such challenges should not prevent the basic research needed to understand the impact that basic genetic variations might have on behaviors or attitudes of interest in large population samples, so as to begin to understand how such predispositions might interact with specific environmental cues and triggers either to mute or exacerbate the expression of particular behavioral tendencies.

A final way in which to examine the impact of genetic and biological influences on political behaviors of interest is through the use of hormonal markers, such as testosterone and cortisol. Such hormones may not decisively influence behavior on their own, but they can potentiate certain behaviors under specific conditions. For example, high levels of testosterone appear to increase the likelihood of responding to threats or challenges in a hostile manner. Anyone who doubts the decisive, unconscious, and often unwelcome influence of such forces need only observe a teenager trying to get dressed, a pregnant woman trying to resist a food craving, or a woman in the midst of a hot flash in menopause (or men buying testosterone replacement in an attempt to recapture the vitality of youth) to appreciate the overwhelming influence of such forces on virtually every aspect of social life.

A third way in which experimental work might help illuminate processes of interest concerns the way in which emotion affects decisions and behaviors. Mood manipulation offers one experimental way in which specific emotions, such as fear or anger, can be induced to examine their impact on specific behaviors, such as cooperation, willingness to compromise, or proclivity to respond aggressively to threat. For example, such studies seem informative in striving to understand the ways in which emotion might serve as a signal of commitment or intention between interested parties. Both verbal and nonverbal cues may influence how leaders seek to manipulate emotion in their attempt to influence various audiences or inspire compromise from both domestic and international opponents in situations involving bargaining and negotiation. Furthermore, emotion appears to be critical in forming the basis and maintenance of processes of social and political identity. Bonds of affiliation promote the formation of social and political groups, just as feelings of enmity can dissolve such associations; understanding such processes clearly provides a key piece for understanding the nature of political action and involvement.

Emotional processes also likely prove essential in attempts at reconciliation following ethnic strife, civil war, or genocide in particular. Experimental manipulations of specific emotions can advance our understanding of how specific emotions influence political and social behaviors of interest.

## **VI CHALLENGES**

Despite the promise offered by experiments to help investigate processes of interest to political scientists, several challenges to their wider utility and applicability remain.

The main intellectual challenge involves the difficulty of designing simple and compelling demonstrations of realistic and important political processes at either the individual or the small-group level because most of these processes occur at higher levels of aggregation that are difficult to manipulate. This is why voting offers such an appealing compromise: it offers the possibility to examine individual behavior whose simple aggregation results in a macro outcome. Most important, coherent and cohesive theoretical hypotheses must motivate and inform empirical investigation. In particular, studies should be designed to test competing theoretical propositions against each other, not simply to engage in “dust bowl” empiricism to determine what emerges from the data after the study is run. Deriving testable hypotheses from dominant models, especially competing models that may predict divergent outcomes, needs to occur prior to the design of particular experimental protocol materials. Such specification allows investigators to know what variables to measure, how to do so, and how to manipulate their introduction in ways that test meaningful propositions.

In addition, clear expectations about which independent variables might produce particular outcomes provide experimenters with direction concerning the design of dependent variables. For effective demonstrations, such variables need to encompass sufficient variation for effects to appear, as well as sufficient magnitude for differences to be witnessed. Nothing is more frustrating than designing a careful experiment that results in ceiling or floor effects because too much attention was paid to operationalizing the independent variable and not enough to constructing a viable dependent measure. Carefully operationalizing variables to suit the issue at hand can prove daunting; making tasks that engage subjects without distracting their attention to extraneous details requires special skill. Findings need to be replicated, extended across domains, and expanded to other populations in order to determine the limits of applicability. Aggregation provides the key to discovering and understanding larger macro-level political processes and understanding the boundaries of applicability, particularly in the realm of cross-cultural or broad-based population studies. Such pursuits are not trivial in nature. Often, technology can entice, improve, and change faster than scholarly ability to assimilate its meaning and importance. However, the topic must always drive the substance of inquiry; meaning should never relegate subservience to method.

Two important intertwined institutional and structural concerns emerge as central challenges to the wider applicability of experimental methods as well. First, to accomplish much of this experimental work successfully in an optimal and efficient manner requires truly interdisciplinary research. An MRI study, for example, will require collaboration with someone who has access to a machine; similarly, hormonal or genetic studies need laboratories in which to collect and analyze samples. At a deeper level, understanding the impact of these micro-level variables on macro factors of political interest often requires collaboration with biologists, geneticists, evolutionary psychologists, or biological anthropologists. Such collaboration not only can enrich and enlighten studies in both theoretical and empirical ways but also challenges individual researchers to read outside their primary field and to often go outside their home departments to locate those who share their interest to find colleagues willing to work with them on particular projects of mutual concern.

Second and related, experiments such as those discussed previously demand effective collaboration. Both interdisciplinarity and collaboration often run contrary to disciplinary norms that historically prize and reward the lone scholar doing mainstream work. This proves particularly problematic in the arena of new technologies because young scholars are most likely to prove interested and adept at learning how to use these techniques, and yet they are the ones most likely to be punished at tenure time when the majority of their publications are coauthored, especially if those coauthors exist outside the discipline. However, it remains almost wholly unrealistic for a political scientist to undertake an MRI or hormonal analysis study entirely on his or her own. On the other hand, the professional costs associated with collaboration mean that a great deal of important work may be neglected in favor of paths of less resistance, if less interest as well. This means not only that the best new work on topics of interest to political scientists may take place outside our discipline without our involvement but also that the work that does get done within the discipline becomes both narrower and shallower as a result of those perverse disciplinary incentive structures.

## VII CONCLUSIONS

Experiments do offer unparalleled ability to ascertain causal inferences from a complex, confusing, and often chaotic world. The ability to isolate factors of interest to determine their influence on outcomes of concern provides one of the most compelling reasons to undertake experimental study. However, in many scholars' views, the messy environment of real-world politics—rife with unintended consequences, blowback effects, and self-conscious actors—renders the applicability and generalizability of such sterile experimental environments limited.

However, the past use of experimental methodology has led to a number of important insights on topics of widespread interest to political scientists, including the impact of certain tactics such as direct mail on voter turnout, the influence of

inter-ethnic identity on trust and cooperation, and the effect of negative advertising on candidate evaluation. No doubt further experimental work remains warranted as new technologies offer unprecedented opportunities to access previously inaccessible aspects of human decision-making processes and render the previously invisible transparent for all to witness in wonder.

## REFERENCES

- Ansolabehere, S., Iyengar, S., & Simon, A. (1999). Replicating experiments using aggregate and survey data: The case of negative advertising and turnout. *American Political Science Review*, 93(4), 901–909.
- Ansolabehere, S., Iyengar, S., Simon, A., & Valentino, N. (1994). Does attack advertising demobilize the electorate? *American Political Science Review*, 88(4), 829–838.
- Arceneaux, K. (2005). Using cluster randomized field experiments to study voting behavior. *The Annals of the American Academy of Political and Social Science*, 601, 169–179.
- Bahry, D., Kosolopov, M., Kozyreva, P., & Wilson, R. (2005). Ethnicity and trust: Evidence from Russia. *American Political Science Review*, 99(4), 521–532.
- Bahry, D., & Wilson, R. (2006). Confusion or fairness in the field? Rejection in the ultimatum game under the strategy method. *Journal of Economic Behavior and Organization*, 60(1), 37–54.
- Bailenson, J., Garland, P., Iyengar, S., & Yee, N. (2006). Transformed facial similarity as a political cue: A preliminary investigation. *Political Psychology*, 27(3), 373–385.
- Beer, F., Healy, A., Sinclair, G., & Bourne, L. (1987). War cues and foreign policy acts. *American Political Science Review*, 81(3), 701–716.
- Beer, F., Sinclair, G., Healy, A., & Bourne, L. (1995). Peace agreement, intractable conflict, escalation trajectory: A psychological laboratory experiment. *International Studies Quarterly*, 39(3), 297–312.
- Burden, B., & Klofstad, C. (2005). Affect and cognition in party identification. *Political Psychology*, 26(6), 869–886.
- Dawes, R., Orbell, J., Simmons, R., & van de Kragt, A. (1986). Organizing groups for collective action. *American Political Science Review*, 80(4), 1171–1185.
- Deutsch, M., Epstein, Y., Canavan, D., & Gumpert, P. (1967). Strategies of inducing cooperation: An experimental study. *Journal of Conflict Resolution*, 11(3), 345–360.
- Druckman, J. (2001a). The implications of framing effects for citizen competence. *Political Behavior*, 23(3), 225–256.
- Druckman, J. (2001b). On the limits of framing effects: Who can frame? *Journal of Politics*, 63(4), 1041–1066.
- Druckman, J. N., Green, D. P., Kuklinski, J. H., & Lupia, A. (Eds.). (2011). *Cambridge handbook of experimental political science*. Cambridge, UK: Cambridge University Press.
- Eberhardt, J., Davies, P., Purdie-Vaughns, V., & Johnson, S. L. (2006). Looking deathworthy: Perceived stereotypicality of black defendants predicts capital-sentencing outcomes. *Psychological Science*, 17(5), 383–386.
- Gerber, A., & Green, D. (2000a). The effect of a nonpartisan get-out-the-vote drive: An experimental study of leafletting. *Journal of Politics*, 62(3), 846–857.
- Gerber, A., & Green, D. (2000b). The effects of canvassing, telephone calls and direct mail on voter turnout: A field experiment. *American Political Science Review*, 94(3), 653–663.
- Gerber, A., Green, D., & Shachar, R. (2003). Voting may be habit-forming: Evidence from a randomized field experiment. *American Journal of Political Science*, 47(3), 540–550.

- Geva, N., Mayhar, J., & Skorick, J. M. (2000). The cognitive calculus of foreign policy decision making: An experimental assessment. *Journal of Conflict Resolution*, 44(4), 447–471.
- Gosnell, H. E. (1926, November). An experiment in the stimulation of voting. *American Political Science Review*, 20(4), 869–874.
- Green, D. (2004). Mobilizing African-American voters using direct mail and commercial phone banks: A field experiment. *Political Research Quarterly*, 57(2), 245–255.
- Hyde, S. D. (2007). The observer effect in international politics: Evidence from a natural experiment. *World Politics*, 60(1), 37.
- Hyde, S. D. (2010). Experimenting in democracy promotion: International observers and the 2004 presidential elections in Indonesia. *Perspectives on Politics*, 8(2), 511–527.
- Iyengar, S. (2002). *Experimental designs for political communication research: From shopping mall to the Internet*. Presented at the Workshop in Mass Media Economics, Department of Political Science, London School of Economics. Accessed 8/9/2006 at, <http://pcl.stanford.edu/common/docs/research/iyengar/2002/expdes2002.pdf>.
- Iyengar, S., Peters, M., & Kinder, D. (1982). Experimental demonstrations of the “not-so-minimal” consequences of television news advertising. *American Political Science Review*, 76(4), 848–858.
- Kuklinski, J., Sniderman, P., Knight, T. P., Tetlock, P., Mellers, G., & Mellers, B. (1997). Racial prejudice and attitudes toward affirmative action. *American Journal of Political Science*, 41, 402–419.
- Lieberman, M., Schreiber, D., & Ochsner, K. (2003). Is political cognition like riding a bicycle? How cognitive neuroscience can inform thinking on political thinking. *Political Psychology*, 24(4), 681–704.
- Lobo, R. A. (2013). Where are we 10 years after the Women’s Health Initiative? *The Journal of Clinical Endocrinology and Metabolism*, 98(5), 1771–1780.
- Maliniak, D., Oakes, A., Peterson, S., & Tierney, M. J. (2011). International relations in the US academy. *International Studies Quarterly*, 55(2), 437–464.
- McDermott, R. (2002). Experimental methods in political science. *Annual Review of Political Science*, 5, 31–61.
- Miguel, E. (2004). Tribe or nation? Nation building and public goods in Kenya and Tanzania. *World Politics*, 56(3), 327–362.
- Mintz, A., & Geva, N. (1993). Why don’t democracies fight each other? An experimental study. *Journal of Conflict Resolution*, 37(3), 484–503.
- Mintz, A., Geva, N., Redd, S., & Carnes, A. (1997). The effect of dynamic and static choice sets on political decision making: An analysis using the decision board platform. *American Political Science Review*, 91, 553–566.
- Ostrom, E. (2002). Some thoughts about shaking things up: Future directions in political science. *Political Science and Politics*, 35(2), 191–192.
- Ostrom, E., Walker, J., & Gardner, R. (1992). Covenants with and without a sword: Self-governance is possible. *American Political Science Review*, 86(2), 404–417.
- Oxley, D. R., Smith, K. B., Alford, J. R., Hibbing, M. V., Miller, J. L., Scalora, M., et al. (2008). Political attitudes vary with physiological traits. *Science*, 321(5896), 1667–1670.
- Peterson, S., Tierney, M., & Maliniak, D. (2005). *Teaching and research practices, views in the discipline, and policy attitudes of international relations faculty at U.S. colleges and universities*. Available at, <http://mjtier.people.wm.edu/intlpolitics/teaching/papers.php>.
- Posner, D. (2004). The political salience of cultural difference: Why Chewas and Tumbukas are allies in Zambia and adversaries in Malawi. *American Political Science Review*, 98(4), 529–545.
- Ramirez, R. (2005). Giving voice to Latino voters: A field experiment on the effectiveness of a national nonpartisan mobilization effort. *The Annals of the American Academy of Political and Social Science*, 601, 66–84.

- Sell, J., Griffith, E., & Wilson, R. (1993). Are women more cooperative than men in social dilemmas? *Social Psychology Quarterly*, 56(3), 211–222.
- Smith, R. (2002). Should we make political science more of a science or more about politics? *Political Science and Politics*, 35(2), 199–201.
- Stanford Magazine. (2006, July/August). Morph, whack, choose.
- Valentino, N., Hutchings, V., & White, I. (2002). Cues that matter: How political ads prime racial attitudes during campaigns. *American Political Science Review*, 96(1), 75–90.
- Van de Kragt, A., Orbell, J., & Dawes, R. (1983). The minimal contributing set as a solution to public goods problems. *American Political Science Review*, 77(1), 112–122.
- Wantchekon, L. (2003). Clientalism and voting behavior: Evidence from a field experiment in Benin. *World Politics*, 55(3), 399–422.
- Weitz-Shapiro, R., & Winters, M. (2010). *Lacking information or condoning corruption? Voter attitudes toward corruption in Brazil*. Paper presented at the annual meeting of the American Political Science Association, Washington, DC.
- Wilson, R., Stevenson, R., & Potts, G. (2006). Brain activity in dominant and mixed strategy games. *Political Psychology*, 27(3), 459–478.
- Wong, J. (2005). Mobilizing Asian American voters: A field experiment. *The Annals of the American Academy of Political and Social Science*, 601, 102–114.

## Chapter 14

# Voting and Agenda Setting in Political Science and Economics

Rick K. Wilson

*Rice University, Houston, Texas*

## I INTRODUCTION

How does a group with diverse interests decide on a single collective choice? For example, how do members of Congress decide on a level of funding for the National Science Foundation? How do members of an academic department decide which subfield will be allowed a new faculty hire? Political scientists and economists have a rich theoretical and experimental tradition addressing these questions. In many ways, this area of study constitutes a mature research program in which the main findings are deductively elaborated and well supported by experimental data. Today, the results are taken for granted and are part of the political science canon.

In the next section, I detail the “standard experiment” used by political scientists and economists when understanding group decision-making. This experiment and the accompanying model serve as the foundation for most studies of collective decision-making. The third section presents findings for “institution-free” settings with and without equilibrium. The fourth section turns toward agenda-setting mechanisms, and the fifth section discusses asymmetric power. The final section concludes with cautionary comments for theorists and experimentalists.

## II THE CANONICAL EXPERIMENT

Suppose a three-person department had to decide what kind of position to hire. The possible choices might be to hire an experimentalist, a statistician, or a theorist. How should the department decide? Of course, if all three members of the department (rightly) agree that an experimentalist is needed, then any one of the members could make the decision and everyone would be happy. However, what happens if there is disagreement over the type of position? Because only one position can be filled, some decision rule must be used in order to make the decision. The department might decide on simple majority rule (in this case,

the “winning” position is determined by two of the three department members). Although this seems “democratic” in that no single individual makes the decision for the group, it turns out that guaranteeing a consistent collective (group) choice is difficult.

To illustrate the nature of this problem, suppose the faculty has the following preferences over an experimentalist [E], a statistician [S], and a theorist [T]:

$$\text{Amy : } E > S > T$$

$$\text{Bob : } S > T > E$$

$$\text{Cathy : } T > E > S$$

Let “ $>$ ” note a preference relation—that is, in Amy’s case, an experimentalist is preferred to a statistician; a statistician is preferred to a theorist; and by transitivity an experimentalist is preferred to a theorist. These preferences make it clear that the department disagrees over what type of person should be hired. So how should the decision be made? Appealing to democratic principles, Bob suggests that majority rule be used and proposes that the department first vote between a theorist and an experimentalist. In such a vote, the theorist wins, with Bob and Cathy voting in favor. Bob then proposes that a vote take place between the theorist and the statistician. Here, the statistician wins, with Bob and Amy voting in favor (both preferring a statistician to an experimentalist). The department’s decision, then, is a statistician. However, Amy recognizes that there has been no vote between an experimentalist and a statistician. If she is allowed to call a vote, then an experimentalist will defeat the statistician, with the department’s choice being an experimentalist. Of course, Cathy would recognize what has happened and call for a vote on the theorist. This could continue ad infinitum with no end to the agenda.

The Marquis de Condorcet recognized this problem with simple majority voting in 1785. Much later, Kenneth Arrow (1963) demonstrated that any decision rule violates basic axioms concerning collective choice and calls into question appeals to democratic fairness when making group decisions. It turns out that Arrow’s findings, which were limited to discrete alternatives, are general. An equilibrium is unlikely, and in its absence “anything can happen.” For those who believe that democratic systems always will reveal the popular will, this is disturbing.

The spatial model of committee voting has served as the standard model for theorists and laboratory experimentalists alike.<sup>1</sup> That model had its clearest statements in the late 1970s in a series of papers running through McKelvey (1976, 1979), Schofield (1978), Cohen (1979), and Cohen and Mathews (1980). It states that equilibrium will be rare if: (1) actors have

---

1. For a useful overview of spatial modeling and its historical antecedents, see Enelow and Hinich 1984). An update with applications is given by Dewan and Shepsle (2011).

well-defined preferences over a multidimensional policy space; (2) no actor has specialized agenda power; (3) simple majority rule is employed; and (4) a forward-moving, binary-comparison, agenda mechanism is used.<sup>2</sup> These results culminated in [Riker's \(1980\)](#) pessimistic conclusion that disequilibrium in collective choice should be pervasive and that political science was proper heir to the title of “the dismal science.” However, it is rare to find natural settings matching such conditions, so confirming or refuting the standard model is difficult.<sup>3</sup>

Of course, as “Amy” might ask, “That’s fine for theory, but we constantly see groups making decisions, so how do these models stack up?” Experimentalists are interested in this question and developed a canonical experimental design in order to test a variety of models. The first of these experiments appeared in [Berl, McKelvey, Ordeshook, and Winer \(1976\)](#), [Fiorina and Plott \(1978\)](#), and [McKelvey, Ordeshook, and Winer \(1978\)](#). I detail an experimental design I developed based on these early articles. The design is representative of all spatial committee experiments and requires knowing about the actors, the choice space, preferences, voting rules, and the agenda.

## A Actors

In the design discussed here, there are 5 committee actors. In other committee experiments, the number of subjects ranges from 3 members to 11. An odd number of committee members is typically used in order to avoid tie votes under simple majority rule. Given that each committee decision usually yields a single observation, increasing the size of the committee is costly and normally not worth it. Moreover, results from 5-member committees generalize to all of the standard theoretical outcomes.

Subjects in my experiments used computers and were separated by partitions so that they could not see one another. Using computers was a deliberate choice and aimed at eliminating a number of threats to internal validity. An early experiment by [Laing and Olmsted \(1978\)](#) noted that groups often make odd choices that could be due to subject communication. [Eavey \(1991\)](#) demonstrates that when subjects talk with one another, they key in on concepts of fairness within the group (but see [Grelak & Koford \(1997\)](#) for another view on this matter). In my design, subjects cannot talk with one another, and the computer handles all communications. As usual, subjects are randomly assigned to conditions. This highly stylized design is very close to the stark world depicted by theorists.

---

2. The conditions required to fit such a model are more stringent than those pointed to here. To save space, the reader is invited to look at any of the papers cited previously. [Cox \(1987\)](#) develops explicit conditions for equilibrium; [Krehbiel \(1988\)](#) reviews the general foundations of the model and empirical findings in legislative settings; and [Ostrom, Gardner, and Walker \(1994\)](#) offer a brief discussion of what would be minimally necessary to describe this setting.

3. [Jillson and Wilson \(1994\)](#) claim that the Continental Congress in the later 1770s resembled just such a “McKelvey world.” I find that study compelling.

## B The Policy Space

What did subjects make decisions about? In political science, the decision space is usually referred to as a “policy space.” In the usual committee experiment, the policy space is nothing more than a plane made up of two orthogonal dimensions labeled X and Y. An outcome (a decision) is simply a point in the space given by a pair of (x,y) coordinates. For example, the dimensions could be thought of as different levels of spending for the military and education. An outcome is a spending bundle for both. The policy space used in the computerized experiment is quite dense. It was made of 300 points on the X dimension and 300 points on the Y dimension, representing 90,000 distinct outcomes that could be chosen. Such a policy space closely approximates theoretical models that make strong assumptions about the mathematical topology of the policy space (see [McKelvey, 1976](#)).

Why were there two dimensions and not one or three? For much of the theoretical literature, one dimension is not of much interest. It (almost) always yields equilibrium and is well understood.<sup>4</sup> Three dimensions, obviously, are difficult to present to subjects. It turns out that a two-dimensional policy space has many of the same characteristics as a many-dimensional policy space.<sup>5</sup> As such, it has become the standard design.

As a key feature of the experimental design, the policy space is kept abstract in order to minimize subjects’ bringing their own values into the experiment. If the policy space instead were represented concretely as some mixture of military and education expenditures (or any other pairing of policies), then subjects might think about the space in ways that the experimenter does not control. For example, I might be predisposed to education and ill-disposed to military spending. Even if the experimenter tries to get me to think otherwise about the space (by telling me that I am a “hawk” in this particular committee decision), I might resist because of my own personal preferences. To keep it abstract, subjects are told they are making a decision over the X and Y grid.

## C Preferences

A crucial feature of the experimental design involves motivating subjects and manipulating what they prefer in the policy space. To do so, each subject is assigned an “ideal point” in the policy space and given a payoff function that details exactly how much every alternative is worth. Their ideal point is worth the most to a subject. Representative payoff functions and ideal points are given in [Table 14.1](#) for several different experiments. Typically, subject payoffs are

---

4. This, of course, is an overstatement. Much of the theoretical and empirical work in political science relies on a single dimension for motivation precisely because there is a well-known equilibrium. References to the “median voter” are rife, and these rely on the existence of a single dimension. [Krehbiel \(1988\)](#) provides a statement of the median voter in legislative settings.

5. The interested reader should see [Schofield \(1985\)](#) or [Saari \(1994\)](#).

**TABLE 14.1** Parameters Used in Experiments

Core Preferences			
Member	Ideal points	Max. value	Loss rate ( $\gamma$ )
1	(120,125)	\$15.00	-0.018
2	(34,168)	\$19.00	-0.013
3	(242,247)	\$25.00	-0.011
4	(222,74)	\$19.00	-0.013
5	(30,35)	\$19.00	-0.013
Status quo=(175,265)			
Star Preferences			
Member	Ideal points	Max. value	Loss rate ( $\gamma$ )
1	(22,214)	\$25.00	-0.013
2	(171,290)	\$25.00	-0.013
3	(279,180)	\$25.00	-0.013
4	(225,43)	\$25.00	-0.013
5	(43,75)	\$25.00	-0.013
Status quo=(280,280)			
Skew Preferences			
Member	Ideal points	Max. value	Loss rate ( $\gamma$ )
1	(75,290)	\$30.00	-0.0129
2	(270,118)	\$20.00	-0.0129
3	(240,43)	\$20.00	-0.0170
4	(195,21)	\$20.00	-0.0129
5	(30,64)	\$25.00	-0.0129
Status quo=(280,280)			

*Utility for any  $\mathbf{X}$  and for the  $i$ th's member's ideal point,  $\mathbf{Xi}$ , is given by the following:*

Nonlinear payoff:  $Ui = (\text{max. value}) \times \exp(\gamma \times (||\mathbf{X} - \mathbf{Xi}||))$

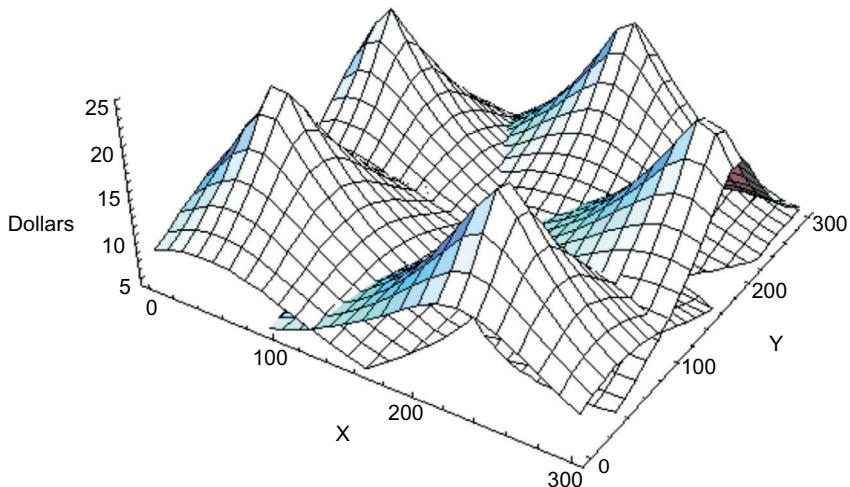
Linear payoff:  $Ui = [(\text{max. value}) - (||\mathbf{X} - \mathbf{Xi}|| \times \$14)]$

represented as circles (indifference curves) with the value of any alternative decreasing as a function of distance from the ideal point.

Subjects are paid in the experiment for the value of the outcome that the group selects. By assigning subjects an ideal point and an associated payoff function, such a design adheres to principles of “induced valuation” promulgated

by [Smith \(1982\)](#). Because I am interested in manipulating the preferences of subjects, I need to ensure they are motivated. In order to do so, there are three concerns. First, the payoff medium must be monotonic in the sense that a subject values more of the reward medium to less. Here, cash, rather than some other reward, is useful. If subjects were rewarded with candy rather than dollars, they might tire of too much candy. With something as fungible as money, each additional increment allows more of something to the subject. The second point is that the reward medium should be salient. Again, the fungibility of money makes it more salient to a subject than many other forms of reward (e.g., additional grade points). Finally, the payoff medium should be dominant. That is, the amount of the reward in the experiment should overcome boredom by the subjects, trump experimenter demand, and exceed the subject's own opportunity costs.

One positive feature of spatial committee experiments is that subject ideal points and payoffs are easily manipulated. If the payoffs meet conditions for “induced valuation,” then subject preferences are a part of the experimental control and variation across outcomes is due to experimental manipulation. [Figure 14.1](#) gives a three-dimensional visual representation of the ideal points and payoff functions for the “star” preference configuration given in [Table 14.1](#). The policy dimensions X and Y are labeled on the figure. The third axis presents the dollar payoffs for subjects across the policy space. This figure illustrates the steepness of the payoff functions for each subject. The maximal amount that a subject earns is the “peak” of the plotted function. The downward “slopes” indicate how fast payoffs decline as a function of distance from each ideal point.



**FIGURE 14.1 Plot of utility functions for five committee members in a spatial committee game.** The X and Y axes constitute the policy space. The “dollars” axis indicates the amount a committee member would receive for each specific (x,y) policy.

Because of the way in which the plot is generated, the decreasing slope for each subject is obscured. The point to the figure is that each subject has a unique payoff function that is mapped onto the policy space.

## D Voting and Agenda Rules

The outcome of a committee decision is a single point in the space. In most of the committee experiments discussed here, a simple majority rule is used in which three of the five committee members must agree on the outcome. The experiments use a forward-moving open-agenda procedure in which proposing alternatives, voting, and adjourning are governed under a modified version of *Robert's Rules of Order*. At the outset of the decision, a status quo is presented to all subjects. That status quo is usually far removed in the policy space from all of the subjects, and as such it is not worth much to anyone. Any subject can place a proposal on the floor. A proposal is a coordinate pair different from the status quo and can appear anywhere in the policy space (including a committee member's own ideal point).

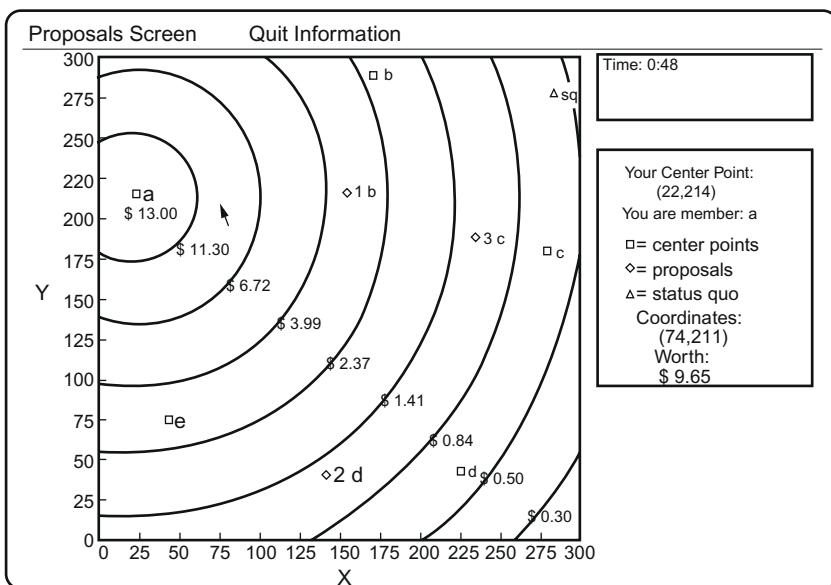
Proposals remain “on the floor” (and on the screen) until seconded by a different person. Once seconded, the proposal is treated as an amendment and a vote is called between it and the status quo. If a majority votes in favor of keeping the status quo, the experiment continues and the floor is opened to further amendments. If a majority votes for the amendment, it becomes the (amended) status quo and the floor is open to further amendments. The experiment continues in this manner until a motion is made to adjourn. Anyone, at any time, can call for adjournment. If a majority votes to adjourn, the decision period ends and subjects are paid in cash the value of the current status quo. If a majority votes against adjournment, the experiment continues, with the floor open to further amendments. It is up to a majority of the committee to decide when to end the decision period.

Everyone in the experiment sees the location of the current status quo, all of the proposals, and the ideal points of the other players; if an amendment is seconded, they see that as well. To make a proposal, a subject uses a pull-down menu and clicks on the point in the screen. The proposal is posted on everyone's screen within 200 milliseconds. Likewise, in order to second a proposal, a subject uses the pull-down menu and clicks on an existing proposal. When a vote is called, the status quo and the amendment flash for 15 seconds, with a warning that a vote is impending. During that time, additional proposals can be put on the floor, although no one can second another proposal. Once the vote is called, the screen changes and subjects see the location and value of the status quo and amendment. Subjects are instructed to vote for one or the other. When everyone finishes voting, the number voting in favor of the status quo and the amendment (but not who voted for each) is reported, and the screen switches back to the policy space. A vote to adjourn has a similar screen, except subjects are told the value of the current status quo and then are asked whether they wish to quit or continue.

## E An Example

Figure 14.2 provides a picture of the main screen used in one of the experiments. The large square box on the left is the policy space for the experiment. This particular screen belonged to member “a,” whose ideal point (the most valuable point to member a) was located at (22,214). A subject’s ideal point is referred to as his or her “center point.” Also plotted for subjects are representative indifference contours (which were called “value circles” in the instructions) and values associated with those contours. In this particular instance, member a’s ideal point was worth \$19, and payoffs declined rapidly as a function of distance away from that ideal point.<sup>6</sup> To be consistent with models of complete information, the ideal points of the other players are displayed, as are any proposals and the current status quo. Also displayed is the location of the cursor. On the figure, an arrow represented the cursor’s position with its tip pointing to (74,211), which is worth \$9.65. The computer automatically calculated the value of that point for the subject. In this way, subjects can obtain much finer readings of value than by trying to interpolate value from the indifference contours.

In this example, the status quo is located in the upper right corner at the point (280,280). Also on the screen are four different proposals. A proposal



**FIGURE 14.2 Sample decision screen for committee member “a.”** The screen provides information concerning payoffs, proposals on the floor, and the ideal points of other members.

6. See Grelak and Koford (1997) for a useful discussion about the steepness of payoff functions and what it means for stability in the distribution of outcomes.

is represented as a diamond and is given a number based on when it was put on the floor. The letter to the right of the number indicates who proposed that alternative. The first proposal, located at (152,218) was made by committee member “*b*.” That member’s ideal point is given by the square and letter located at (171,290). This particular configuration of preferences is equivalent to the “star” configuration given in [Table 14.1](#).

### III EQUILIBRIUM AND DISEQUILIBRIUM

The first spatial committee experiments turned to the question of equilibrium and addressed two clear predictions:

- P1. The equilibrium will be chosen when it exists.
- P2. In the absence of equilibrium, anything can happen.

[Plott \(1967\)](#) and, later, [Cox \(1987\)](#) axiomatically prove the difficulty of ensuring an equilibrium in a multidimensional space. For a five-person committee in a two-dimensional space, it means that four of the committee members must be pairwise symmetrically dispersed around the fifth member.<sup>7</sup> Although there are many ways in which preferences can be arranged to yield equilibrium, there are far more ways in which no equilibrium will occur. The first prediction is a point prediction. When an equilibrium exists, it falls at a single point in the policy space. This means that if an experimenter manipulates the ideal points of subjects so as to guarantee equilibrium, it provides an explicit prediction.

The second prediction is more vexing and has driven most of the theoretical and experimental literature on voting. If there is no equilibrium, then there should be no pattern to the group choice. These predictions are difficult to test in a natural setting because preferences cannot be accurately measured or manipulated. However, spatial committee experiments allow for direct tests.

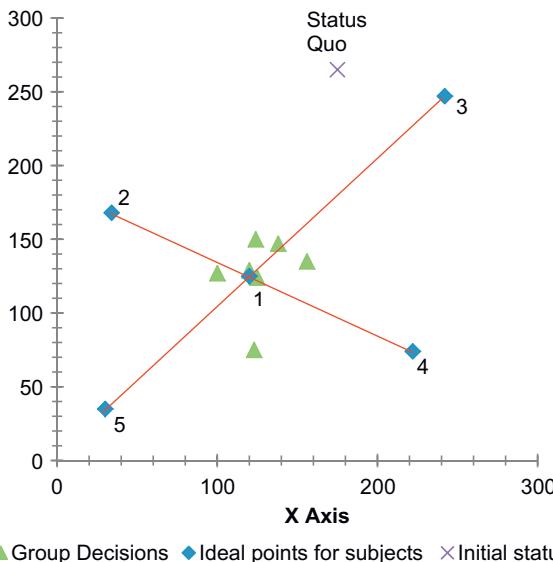
#### A Equilibrium Outcomes

My own results from an experiment with equilibrium are representative and troubling. Seven trials are run that manipulate preferences so that a unique equilibrium exists. The group decisions are plotted in [Figure 14.3](#), as are the ideal points of the five actors and the initial status quo.<sup>8</sup> The unique equilibrium is located at committee member 1’s ideal point. The lines connecting members 2 and 4 and 3 and 5 illustrate that these subjects are pairwise symmetric around 1. From visual inspection, it is easy to see that no outcome is located at the

---

7. There are additional, very strong, assumptions that must be made about the shape of the indifference curves of all of the committee members, among other features of the model. For a readable introduction to spatial models, see [Enelow and Hinich \(1984\)](#).

8. It is important to note that in various trials, players’ ideal points were rotated so that the equilibrium was located in different parts of the alternative space. All of the outcomes have been normalized to the same distribution of ideal points.



**FIGURE 14.3 Group decisions under experiments with equilibrium.** The equilibrium is located at the ideal point for committee member 1.

equilibrium. The equilibrium does not do well as a point prediction, and this finding is true for a number of spatial committee experiments. Berl et al. (1976), Fiorina and Plott (1978), Wilson (1986), Eavey (1991), and Grelak and Koford (1997) all find that outcomes consistently deviate from the equilibrium. At best, these outcomes are “close.”

Fortunately, spatial theories predict more than an outcome. An additional prediction holds that the equilibrium is attractive, which implies that successful amendments will converge to it. In the experiment, data are collected concerning the proposals that are made, the time at which those proposals are made, and all of the votes on amending and adjourning. This allows me to reconstruct the agenda and to test whether successive amendments are closer to the equilibrium than to the status quo they replace. In five of the seven trials, *every* successful amendment was closer to the equilibrium than its predecessor. This strictly satisfies the “attractiveness” component of the model because the agenda converged toward the equilibrium. Moreover, the process converged quickly.

As can be seen from Table 14.2, subjects typically took less than 10 minutes of floor discussion to end the period.<sup>9</sup> On average, the final outcome was selected within 41.6 seconds of beginning the trial. For the five trials converging on the equilibrium, subjects averaged 1.4 amendments before adjourning. This meant that the final outcome was quickly selected, was chosen from a handful

9. The time that elapsed while taking a vote was excluded.

**TABLE 14.2** Descriptive Data for the Baseline Core Trials

Trial	Total Time (Seconds)	Final Outcome	No. of Proposals	No. of Amendment Votes	No. of Adjournment Votes
BCORE1	610	(125,124)	39	14	12
BCORE2	156	(123,75)	17	3	3
BCORE3	548	(156,135)	31	8	8
BCORE4	84	(138,147)	6	2	1
BCORE5	166	(124,150)	17	1	5
BCORE6	1972	(100,127)	40	62	25
BCORE7	398	(120,129)	39	9	4

of proposals on the floor, and was reached via a short agenda. However, this did not mean that subjects were quick to end the trial. The number of unsuccessful amendments outpaced successful amendments by almost 4 to 1, and subjects averaged 7.3 unsuccessful adjournment votes before ending the trial.

At first glance, the findings seem odd. On the one hand, the equilibrium is never chosen; on the other hand, it is attractive. Why this discrepancy? First, only a small number of proposals made it to the floor (on average, 27). Moreover, the equilibrium was *never* proposed in *any* of the experimental trials. Second, subjects tended to make proposals near their own ideal points, but they had to depend on someone else bringing them to a vote. Proposals located at some distance from one's own ideal point were not financially worthwhile, and this led to a limited set of proposals being brought to a vote. The standard spatial model with an equilibrium enjoys more support than it might seem at first glance.

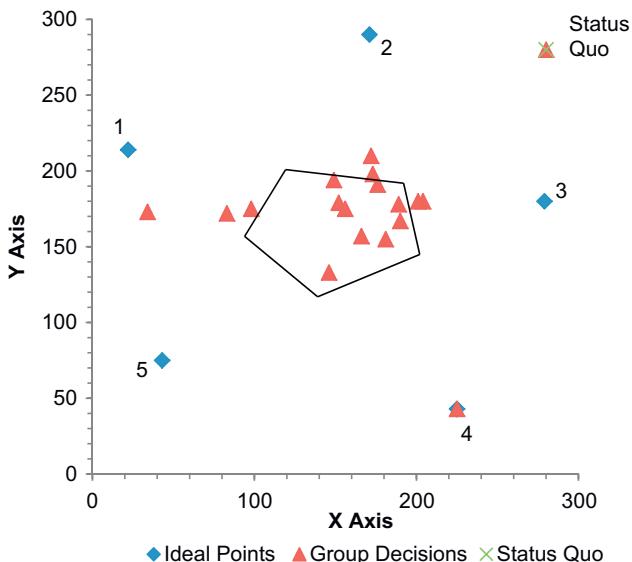
## B Nonequilibrium Outcomes

The more interesting case involves group decisions when there is no equilibrium. This setting is expected to be more common and is all the more troubling because there is no useful prediction.<sup>10</sup> Anything can happen in this case.

I examine two different manipulations of ideal points: a “star” and a “skew star” distribution. In both instances, there is no equilibrium. Under the star distribution, subjects’ ideal points are scattered about the policy space and

---

10. A number of different solution concepts have been proposed and tested. For different approaches, see, for example, Fiorina and Plott (1978); Feld, Grofman, Hartlet, Kilgour, and Miller (1987); Bianco, Lynch, Miller, and Sened (2006); and Godfrey, Grofman, and Feld (2011).



**FIGURE 14.4** Group decisions under disequilibrium “star” preferences.

no one occupies a central position. Given the relative symmetry of the ideal points, there is no obvious minimum winning coalition, nor is there any obvious focal point (Figure 14.4). The skew star manipulates preferences so that three committee members are near one another and form a natural winning coalition.

The trials under the star manipulation are conducted in the same manner as those with the equilibrium manipulation. Outcomes from 18 trials are plotted on Figure 14.4. Once again, the ideal points of the five committee members are displayed, as is the status quo from which each agenda begins. The axiomatic models provide little insight into the distribution of outcomes. The usual view is that because anything can happen, outcomes should be broadly distributed across the policy space. Instead, it is easy to see that outcomes are concentrated in the middle of the space. Half (9 of 18) of the outcomes are located in the (small) interior pentagon made up by finding the convex hulls of all possible winning coalitions. That pentagon defines a central portion for the alternative space. Although three outcomes are located well outside that central space—one at the status quo where a majority voted to end the experiment unusually quickly—most are centrally distributed. The same finding is reported by [Fiorina and Plott \(1978\)](#) and [McKelvey and Ordeshook \(1990\)](#) in their survey of spatial committee experiments.

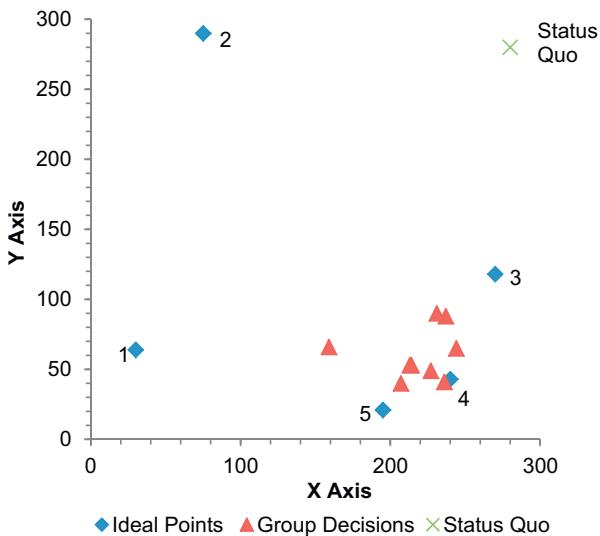
There is an additional empirical implication for disequilibrium: these agendas should be longer and more extensive than when an equilibrium exists. This implication is strongly supported by the data reported in Table 14.3. First, I find that subjects spend more time in these trials: subjects averaged slightly

**TABLE 14.3** Descriptive Data for Baseline Star Trials

Trial	Total Time (Seconds)	Final Outcome	No. of Proposals	No. of Amendment Votes	No. of Adjournment Votes
BSTAR1	951	(176,191)	76	24	3
BSTAR2	77	(225,43)	6	1	1
BSTAR3	712	(204,180)	71	24	9
BSTAR4	782	(172,210)	48	15	6
BSTAR5	1,901	(190,167)	174	55	18
BSTAR6	422	(34,173)	31	10	2
BSTAR7	1,474	(146,133)	115	38	26
BSTAR8	1,377	(173,198)	72	30	20
BSTAR9	39	(280,280)	5	0	1
BSTAR10	1,905	(83,172)	96	37	14
BSTAR11	1,845	(152,179)	72	26	14
BSTAR12	921	(201,180)	41	14	9
BSTAR13	477	(149,194)	39	9	3
BSTAR14	292	(156,175)	19	4	2
BSTAR15	1,376	(98,175)	92	24	1
BSTAR16	1,302	(166,157)	70	21	4
BSTAR17	961	(189,178)	34	20	14
BSTAR18	1,097	(181,155)	61	20	6

more than 15 minutes per trial compared with an average of slightly less than 10 minutes in equilibrium trials. Subjects also make more proposals and do so at a rate that is two and one-half times their counterparts in the equilibrium manipulation. In the most extreme case, subjects in the trial BSTAR5 had 174 proposals on the floor. By comparison, no more than 40 were placed on the floor in any equilibrium trial. Subjects under the star configuration also called a significantly larger number of amendment votes than did their counterparts under the equilibrium configuration (on average, slightly less than 21 for the former compared with slightly more than 14 for the latter).

These findings for the star baseline are concentrated in the central portion of the space. Is there something about that region, or is it simply an artifact of



**FIGURE 14.5** Group decisions under disequilibrium “skew star” preferences.

the preference configuration used? In order to address this question, the ideal points for subjects were manipulated while keeping the structure of the institution the same.

The skew preference configuration, like the star configuration, has no majority rule equilibrium. Three players have ideal points located in the same quadrant of the alternative space, with two other players located some distance away (see the distribution in Figure 14.5). This preference configuration allows us to determine whether the central portion of the policy space has an independent effect on outcomes.

As with the star configuration, these committee choices are clustered in a specific region of the alternative space, but they are removed from the central part of the space. Consequently, it appears there is nothing special about the central region. Two points are important when eyeballing the distribution of these outcomes. First, it is remarkable that outcomes shift so markedly with a change in the distribution of preferences. Clearly, subjects are responding to the manipulation. Second, and more critically, these outcomes are not widely distributed. Once again, I find that outcomes are tightly clustered and are not widely distributed across the policy space. This echoes findings by [Laing and Olmsted \(1978\)](#) and [McKelvey et al. \(1978\)](#), who examined a variety of distributions of ideal points and observed a similar phenomenon.

Even though outcomes are clustered, this did not mean that subjects had an easy time selecting them. Again, I find that the agenda process is lengthy and extensive. First, under the skew star configuration, subjects took a good deal of time to settle on an outcome (Table 14.4). On average, they spent approximately 12.5 minutes in proposal making—less than under the star but

**TABLE 14.4** Descriptive Data for Baseline Skew Star Trials

Trial	Total Time (Seconds)	Final Outcome	No. of Proposals	No. of Amendment Votes	No. of Adjournment votes
BSKEW1	317	(227,49)	32	8	6
BSKEW2	358	(237,88)	32	6	6
BSKEW3	478	(207,40)	36	12	6
BSKEW4	147	(244,65)	32	3	3
BSKEW5	105	(214,53)	8	2	1
BSKEW6	2,346	(213,53)	137	76	21
BSKEW7	415	(231,90)	30	10	6
BSKEW8	474	(236,41)	27	9	8
BSKEW9	1,998	(159,66)	123	46	29

more than under the core configuration. Although subjects in the skew star manipulation cast almost as many amendment votes as subjects in the star configuration, they were less successful in finding amendments to the status quo. On average, they replaced the status quo only 4.0 times versus 6.7 times in the star trials. In part, this was due to the fact that fewer proposals were made that could overturn the status quo. Only 13.7% of the proposals on the floor could have defeated the final status quo, compared with 20.8% for the star replications.

The skew star trials ended more quickly than the star trials. Agendas were typically shorter, and a specific coalition dominated when choosing an outcome. Outcomes were not scattered throughout the alternative space, but neither were they concentrated in the center of that space. Both of the disequilibrium manipulations raise interesting questions about standard spatial models. On the one hand, the experiments show that outcomes converge. On the other hand, the agendas for many of these trials demonstrate the kind of incoherence that theorists expect.

## IV AGENDAS

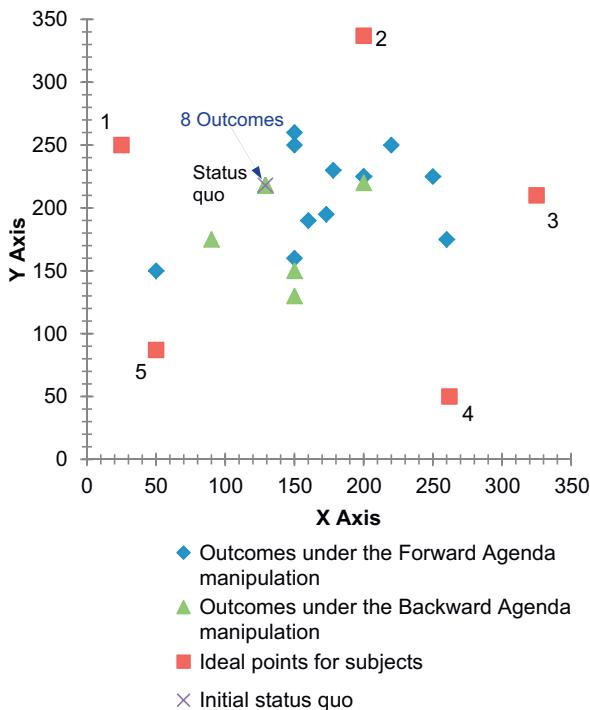
The absence of equilibrium poses a serious challenge for scholars. A good deal of theoretical work focuses on institutional mechanisms that impose equilibrium (see Diermeier & Krehbiel, 2003; Shepsle, 1989). Control over the agenda is an important focus for political scientists. Romer and Rosenthal (1978) and

[Shepsle \(1979\)](#) provide the earliest statements of the ways in which agenda mechanisms directly affect outcomes in settings in the absence of equilibrium. The former shows the advantage held by those close to the status quo, whereas the latter points to germaneness rules in legislative settings that yield equilibrium.

[Plott and Levine \(1978\)](#) illustrate agenda power both in a natural setting (an airplane club) and in an experimental setting. When the preferences of others are well-known and when control over the agenda is ceded to an agenda setter, then predicting the group decision is easy: it will be the agenda setter's ideal point. [Kormendi and Plott \(1982\)](#) and [Endersby \(1993\)](#) point to a variety of agenda rules and their impact on outcomes. Of course, as [Eavey and Miller \(1995\)](#) show, agenda setting may not always matter if agenda setters have extreme preferences. [Krehbiel \(1986\)](#) makes a similar point in a legislative context. Generally, there is consensus that agenda rules yield equilibrium, and there is considerable experimental support for such a position (see reviews by [McKelvey & Ordeshook \(1990\)](#), [Palfrey \(2009\)](#), and [Miller \(2011\)](#)).

There are many variations on agenda rules, and here I call on an example from my own research. In many deliberative bodies, a “backward-moving” agenda is used. Unlike the forward-moving agenda described in the previous section, in this setting the status quo is voted last. The agenda is built so that the last amendment is voted first and is treated as a perfecting amendment to the prior amendment. The last amendment standing is voted against the status quo in the final vote. Suppose there is an agenda that looks like {sq, a, b, c, d}, where the letters represent amendments and their order. The first vote is between “c” and “d.” The winner of that vote is then paired with “b” and so forth. In the final vote, the winning amendment is paired with “sq,” the status quo. [Shepsle and Weingast \(1984\)](#) develop the equilibrium for this procedural rule, and the experiment I designed tests it directly.

In this experiment, a star preference configuration is used. Rather than beginning with the status quo in the upper right corner of the policy space, a status quo was chosen that minimized the size of the equilibrium under a backward-moving agenda (no equilibrium exists under a forward-moving agenda). The status quo was the same for both the backward- and the forward-moving agenda. [Figure 14.6](#) plots the committee outcomes for both manipulations. Although it may appear that there is no difference between the two manipulations because of the distribution of outcomes, the differences are dramatic. Outcomes under the forward-moving agenda range across the policy space in much the same manner as noted in the prior section. By comparison, all of the outcomes under the backward-moving agenda are in equilibrium. Eight of the 12 committee decisions are located at the status quo, indicating its advantage under this institutional mechanism. None of the outcomes from the forward-moving agenda returned to the initial status quo. The conclusion from this experiment is that agendas matter.



**FIGURE 14.6** Group decisions with forward and backward agenda manipulations.

The earliest findings from spatial committee experiments turned on questions of the agenda procedures that induce equilibrium. For many, it was refreshing to discover that institutional design could ensure equilibrium and that the equilibrium could be predicted. The question next raised was what other types of institutional rules affect group choices?

## V ASYMMETRIC RELATIONS

Political scientists and economists naturally understood that the symmetric power relations usually assumed in spatial committee theoretical models were unrealistic. After all, political institutions typically cede power to some at the expense of others. An early discussion by Buchanan and Tullock (1962) pointed to the trade-offs when different-sized aggregation rules are used. For example, the advantage that a person holds under unanimity is very different than when a simple majority is needed. Theorists such as Schofield (1985) spent a good deal of time demonstrating the circumstances in which equilibrium emerges as the size of the vote and the dimensionality of the policy space changes. The possibility that a single player could be given the power to veto was discussed by Slutsky (1979); later, Tsebelis (2002) pointed to how institutions can be

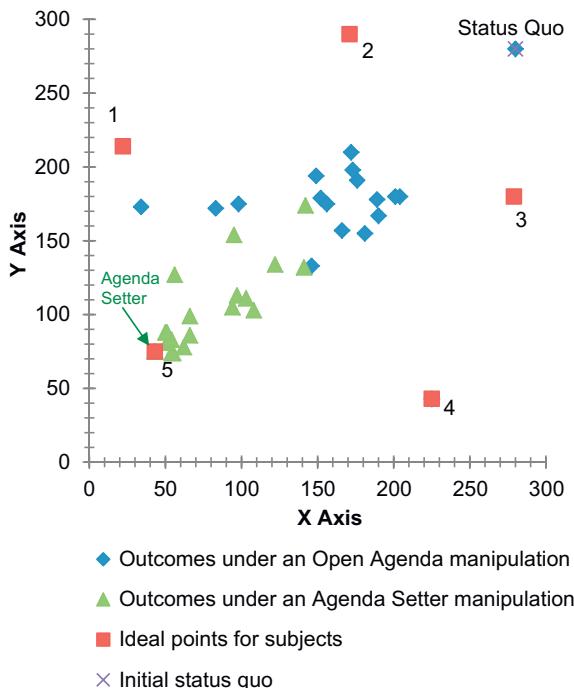
designed to effect veto powers. Others pointed to equilibrium appearing through multiple institutions, noting that a balance between the U.S. House and Senate effectively provides vetoes by each branch (Miller, 1987). Finally, Baron and Ferejohn (1989) developed a model illustrating the power of the initial proposer in legislative settings. This is only a sampling of the work that sorts out what happens when some actors are given special powers.

There is also a rich set of experiments illustrating what happens as special rules are granted to some players. Experiments by Laing, Nakabayashi, and Slotnick (1983) and King (1994) demonstrate the effect of changing aggregation rules. Not surprisingly, building an agenda gets more difficult as the size of the majority increases. When unanimity is required, committee decisions are virtually deadlocked. Wilson and Herzberg (1987) and Haney, Herzberg, and Wilson (1992) examined different versions of vetoing and blocking decisions. Any actor assigned veto power is greatly advantaged, and committee decisions reflect this fact. Miller, Hammond, and Kile, (1996) and Bottom, Eavey, Miller, and Victor (2000) provide evidence noting the balance provided by institutions when committee members must decide first within a group and then across groups. This simple change in the decision setting yields an equilibrium, and subjects converge on it. Bottom, Eavey, and Miller (1996) illustrate the power of the initial proposer in spatial committee games, although they found that such power is not absolute. By and large, the experimental data support the theoretical models, noting that when some are given power over others, the group decisions reflect that advantage.

What does it mean to give a committee member institutional advantage? To illustrate this, I again use my own research. In this experiment, one subject was randomly chosen and granted the sole power to bring a proposal to a vote and to call a vote to adjourn. Other committee members kept the power to place proposals on the policy space, but they could not second any proposals, nor could they call for a vote to adjourn. In effect, monopoly power is assigned to a single individual. Other aspects of a standard committee experiment remain the same. A vote to amend the status quo still requires a simple majority and so too does adjournment.

Figure 14.7 illustrates the outcomes from a standard committee experiment using a forward-moving agenda and outcomes from the monopoly agenda-setter experiment. The subject at member 5's position was *always* assigned to be the agenda setter, and this point was also the equilibrium for the manipulation. These outcomes are represented as diamonds, and outcomes under the standard manipulation are represented as circles. Two points are clear from the figure. First, although the structure of preference is the same for both manipulations, the pattern of outcomes is quite different. Under the standard manipulation, outcomes are more scattered in the policy space. Outcomes under the agenda-setter manipulation are more compact, and they are anchored to the agenda setter's ideal point.

Second, group decisions under the agenda-setting manipulation do not fall at the equilibrium. Instead, those outcomes range from the agenda setter's



**FIGURE 14.7** Group decisions with specialized agenda and open-agenda manipulations.

ideal point to the central portion of the alternative space. This indicates that the agenda setter's power is not total, although that individual is advantaged by having the right to call votes. What is valuable about this experiment is that it points to simple ways in which rules can be changed, how it is easy to instantiate those changes, and how theory and experiment can work together. It also makes clear that there remains a gap between theory and empirical outcomes.

## VI EQUILIBRIUM AND DISEQUILIBRIUM, REDUX

I have sampled only a few of the results from several decades of work on group decision-making in political science and economics. This work is rather stylized and focuses on institutional mechanisms that yield equilibrium. Much of this work was in response to theoretical models predicting disequilibrium. Experimental results on agenda setting and asymmetric power reoriented political scientists and economists to understanding basic institutional mechanisms.

The standard axiomatic model for social choice continues to present an intriguing empirical problem. As I briefly noted in [Section III](#), group decisions do not necessarily converge on the equilibrium. As well, group decisions do not wander throughout the policy space when there is no equilibrium; in fact, in these settings the outcomes tend to cluster in distinct regions and respond to

a shift in committee members' ideal points. As such, it seems there is room for continued theorizing and making sense of the empirical findings.

Several directions have been taken when confronting these anomalous findings. Among the axiomatic theorists, [Miller \(1980\)](#) shows that with finite alternatives, the collective choice will not cycle among all alternatives when there is no unique equilibrium. [McKelvey \(1986\)](#) extends Miller's conjectures to multi-dimensional spaces and proves that outcomes will be limited to a distinct subset of the alternative space that has been coined the "uncovered" set. [Bianco et al. \(2006\)](#) have gone back through many of the spatial committee experiments and calculated the uncovered set for each distribution of ideal points. This has been no mean feat because, although the theoretical properties of this set are well-known, it is only with very fast computers that it has been possible to identify this set. [Bianco et al. \(2006\)](#) find that while disequilibrium remains the rule of the day, at least outcomes are limited to a subset of the policy space. Much of this work is driven by strictly rational models of human behavior.

I think the more worthwhile path to take is to join with others who call for behavioral models that inform theoretical models. [Camerer \(2003\)](#) clearly makes this point when examining bargaining and negotiation experiments in economics. Behavioral biases, simple heuristics, and cognitive constraints may all play a part for informing our basic models. It seems obvious that people do not finely search a complex policy space, that people are unwilling to invest considerable time pondering a cognitively demanding task, and that patience may wear thin. Fairness concerns or even envy may intervene when making choices. In a broad survey of the literature, I note several efforts by political scientists in exploring these behaviors ([Wilson, 2010](#)). By understanding behavior in the laboratory, we can better inform our axiomatic models. In turn, those models can open up new avenues for research and insight into collective choice.

## ACKNOWLEDGMENTS

I gratefully acknowledge the support of the Workshop in Political Theory and Policy Analysis and the National Science Foundation (SES 8721250). Neither bears responsibility for the conclusions reached in this chapter.

## REFERENCES

- Arrow, K. J. (1963). *Social choice and individual values*. New Haven, CT: Yale University Press.
- Baron, D. P., & Ferejohn, J. A. (1989). Bargaining in legislatures. *American Political Science Review*, 83(4), 1181–1206.
- Berl, J. E., McKelvey, R. D., Ordeshook, P. C., & Winer, M. D. (1976). An experimental test of the core in a simple n-person cooperative non-side payment game. *Journal of Conflict Resolution*, 20(3), 453–476.
- Bianco, W. T., Lynch, M. S., Miller, G. J., & Sened, I. (2006). "A theory waiting to be discovered and used": A reanalysis of canonical experiments on majority rule decision making. *Journal of Politics*, 68(4), 837–850.

- Bottom, W. P., Eavey, C. L., & Miller, G. J. (1996). Getting to the core—Coalitional integrity as a constraint on the power of agenda setters. *Journal of Conflict Resolution*, 40(2), 298–319.
- Bottom, W. P., Eavey, C. L., Miller, G., & Victor, J. N. (2000). The institutional effect on majority rule instability: Bicameralism in spatial policy decisions. *American Journal of Political Science*, 44(3), 523–540.
- Buchanan, J., & Tullock, G. (1962). *Calculus of consent*. Ann Arbor, MI: University of Michigan Press.
- Camerer, C. F. (2003). Behavioral game theory: Experiments in strategic interaction. In C. F. Camerer, & E. Fehr (Eds.), *The roundtable series in behavioral economics*. New York: Russell Sage Foundation/Princeton University Press.
- Cohen, L. (1979). Cyclic sets in multidimensional voting models. *Journal of Economic Theory*, 20(1), 1–12.
- Cohen, L., & Mathews, S. (1980). Constrained Plott equilibrium, directional equilibria and global cycling sets. *Review of Economic Studies*, 97, 975–986.
- Cox, G. (1987). The uncovered set and the core. *American Journal of Political Science*, 31, 408–422.
- Dewan, T., & Shepsle, K. A. (2011). Political economy models of elections. *Annual Review of Political Science*, 14, 311–330.
- Diermeier, D., & Krehbiel, K. (2003). Institutionalism as a methodology. *Journal of Theoretical Politics*, 15(2), 123–144.
- Eavey, C. L. (1991). Patterns of distribution in spatial games. *Rationality and Society*, 3, 450–474.
- Eavey, C. L., & Miller, G. J. (1995). Subcommittee agenda control. *Journal of Theoretical Politics*, 7(2), 125–156.
- Endersby, J. W. (1993). Rules of method and rules of conduct: An experimental study on two types of procedure and committee behavior. *Journal of Politics*, 55(1), 218–236.
- Enelow, J. M., & Hinich, M. J. (1984). *The spatial theory of voting*. Cambridge, UK: Cambridge University Press.
- Feld, S., Grofman, L. B., Hartlet, R., Kilgour, M., & Miller, N. (1987). The uncovered set in spatial voting games. *Theory and Decision*, 23, 129–155.
- Fiorina, M. P., & Plott, C. R. (1978). Committee decisions under majority rule: An experimental study. *American Political Science Review*, 72(2), 575–598.
- Godfrey, J., Grofman, B., & Feld, S. L. (2011). Applications of Shapley–Owen values and the spatial Copeland winner. *Political Analysis*, 19(3), 306–324.
- Grelak, E., & Koford, K. (1997). A re-examination of the Fiorina–Plott and Eavey voting experiments: How much do cardinal payoffs influence outcomes? *Journal of Economic Behavior & Organization*, 32(4), 571–589.
- Haney, P., Herzberg, R., & Wilson, R. K. (1992). Advice and consent: Unitary actors, advisory models and experimental tests. *Journal of Conflict Resolution*, 36(4), 603–633.
- Jillson, C. C., & Wilson, R. K. (1994). *Congressional dynamics: Structure, coordination and choice in the first American Congress, 1774–1789*. Palo Alto, CA: Stanford University Press.
- King, R. R. (1994). An experimental investigation of super majority voting rules—Implications for the financial accounting standards board. *Journal of Economic Behavior & Organization*, 25(2), 197–217.
- Kormendi, R. C., & Plott, C. R. (1982). Committee decisions under alternative procedural rules. *Journal of Economic Behavior & Organization*, 3(3), 175–195.
- Krehbiel, K. (1986). Sophisticated and myopic behavior in legislative committees: An experimental study. *American Journal of Political Science*, 30(3), 542–561.
- Krehbiel, K. (1988). Spatial models of legislative choice. *Legislative Studies Quarterly*, 13(3), 259–319.

- Laing, J. D., Nakabayashi, N., & Slotnick, B. (1983). Winners, blockers, and the status quo: Simple collective decision games and the core. *Public Choice*, 40(3), 263–279.
- Laing, J. D., & Olmsted, S. (1978). An experimental and game theoretic study of committees. In P. C. Ordeshook (Ed.), *Game theory and political science*. New York: New York University Press.
- McKelvey, R. D. (1976). Intransitivities in multidimensional voting models and some implications for agenda control. *Journal of Economic Theory*, 12(3), 472–482.
- McKelvey, R. D. (1979). General conditions for global intransitivities in formal voting models. *Econometrica*, 47(5), 1085–1111.
- McKelvey, R. D. (1986). Covering, dominance, and institution free properties of social choice. *American Journal of Political Science*, 30(2), 283–314.
- McKelvey, R. D., & Ordeshook, P. C. (1990). A decade of experimental research on spatial models of elections and committees. In J. M. Enelow, & M. J. Hinich (Eds.), *Advances in the spatial theory of voting*. Cambridge, UK: Cambridge University Press.
- McKelvey, R. D., Ordeshook, P. C., & Winer, M. D. (1978). The competitive solution for  $N$ -person games without transferable utility with an application to competitive games. *American Political Science Review*, 72(2), 599–615.
- Miller, N. R. (1980). A new solution set for tournaments and majority voting: Further graph-theoretical approaches to the theory of voting. *American Journal of Political Science*, 24(1), 68–96.
- Miller, G. J. (1987). Core of the Constitution. *American Political Science Review*, 81(4), 1155–1174.
- Miller, G. J. (2011). Legislative voting and cycling. In J. N. Druckman, D. P. Green, J. H. Kuklinski, & A. Lupia (Eds.), *Cambridge handbook of experimental political science* (pp. 353–368). New York: Cambridge University Press.
- Miller, G. J., Hammond, T. H., & Kile, C. (1996). Bicameralism and the core: An experimental test. *Legislative Studies Quarterly*, 21(1), 83–103.
- Ostrom, E., Gardner, R., & Walker, J. (1994). *Rules, games and common-pool resources*. Ann Arbor, MI: University of Michigan Press.
- Palfrey, T. R. (2009). Laboratory experiments in political economy. *Annual Review of Political Science*, 12, 379–388.
- Plott, C. R. (1967). A notion of equilibrium and its possibility under majority rule. *American Economic Review*, 57(3), 787–806.
- Plott, C. R., & Levine, M. E. (1978). A model of agenda influence on committee decisions. *American Economic Review*, 68(1), 146–160.
- Riker, W. H. (1980). Implications from the disequilibrium of majority rule for the study of institutions. *American Political Science Review*, 74(2), 432–446.
- Romer, T., & Rosenthal, H. (1978). Political resource allocation, controlled agendas, and the status quo. *Public Choice*, 33(1), 27–43.
- Saari, D. (1994). *Geometry of voting* (Series in Economic Theory Vol. 3). Berlin: Springer-Verlag.
- Schofield, N. (1978). Instability of simple dynamic games. *Review of Economic Studies*, 45(141), 575–594.
- Schofield, N. (1985). *Social choice and democracy*. Heidelberg: Springer-Verlag.
- Shepsle, K. A. (1979). Institutional arrangements and equilibrium in multidimensional voting models. *American Journal of Political Science*, 23(1), 27–59.
- Shepsle, K. A. (1989). Studying institutions: Some lessons from the rational choice approach. *Journal of Theoretical Politics*, 1(2), 131–149.
- Shepsle, K. A., & Weingast, B. R. (1984). Uncovered sets and sophisticated voting outcomes with implications for agenda institutions. *American Journal of Political Science*, 28(1), 49–74.

- Slutsky, S. (1979). Equilibrium under a majority voting. *Econometrica*, 46(5), 1113–1126.
- Smith, V. L. (1982). Macroeconomic systems as an experimental science. *American Economic Review*, 72(5), 923–955.
- Tsebelis, G. (2002). *Veto players: How political institutions work*. New York: Russell Sage Foundation.
- Wilson, R. K. (1986). Results on the Condorcet winner: A committee experiment on time constraints. *Simulations and Games*, 17(2), 217–243.
- Wilson, R. K. (2010). The contribution of behavioral economics to political science. *Annual Review of Political Science*, 14, 201–223.
- Wilson, R. K., & Herzberg, R. Q. (1987). Negative decision powers and institutional equilibrium: Experiments on blocking coalitions. *Western Political Quarterly*, 40(4), 593–609.

## Chapter 15

# Economic Games for Social Scientists

Catherine Eckel

*Texas A&M University, College Station, Texas*

## I INTRODUCTION

Experimental economics is a relatively new field, with the first academic papers published in the 1960s. Only in the past few years has experimental economics become so widely accepted that virtually every economics department believes that it should have an experimentalist on the faculty and train its students in the methodology. In this section, I outline some of the contributions of experimental economics as it has matured, and I sketch the development of the professional infrastructure in the field.

The earliest experimental papers focused almost exclusively on markets (e.g., Smith, 1962), with particular emphasis on finding the conditions under which a market would converge to the competitive price and quantity. Meanwhile, on parallel tracks, psychologists and game theorists began to investigate simple games such as the prisoner's dilemma, and decision theorists tackled individual decision-making under risk and uncertainty, including (but not limited to) expected utility theory. (See Roth (1993) for a discussion of experiments before 1960.)

More than 20 years later, in 1986, as the field really began to take off and gain widespread readership, the first U.S. professional organization of experimental economists, the Economic Science Association (ESA), was formalized by a group of pioneer experimentalists who had been meeting at the Westward Look Resort in Tucson, Arizona, for several years. (The association continues to hold annual meetings at the resort, more than 25 years later.) It was a small group: everyone could fit into a single photograph taken on the hotel's stairs. Vernon Smith was the first president, followed by Charlie Plott.

When I first attended the ESA meetings in 1992, the association was somewhat larger, with approximately 80 papers presented during the 2 days. The program was dominated by two types of experiments: markets and public goods. In the market sessions, auctions and asset markets figured prominently, whereas

in the public goods sessions the word was out: people apparently were much more cooperative than a straightforward extrapolation from simple game theory would predict, and everyone was asking “why” in one way or another. There were also quite a few sessions on collective choice and committee decision-making; this work seems to have nearly died out, suggesting it may be time for a resurgence. Also, the bargaining games that have come to dominate recent meetings had only a small presence—two sessions as I remember—including my own paper on ultimatum bargaining.<sup>1</sup>

Since the early 1990s, a long series of experiments has established regular patterns of violation of standard game theory and expected utility theory. This has led to a remarkable growth of both experimental and theoretical research concerned with how people make real decisions when rationality and information processing are less than perfect and when social considerations such as fairness and cooperation play an important role. I do not mean to suggest that experimentalists have restricted themselves to worrying about fairness and bounded rationality (although for me these are some of the most engaging elements of the experimental agenda).

Experimentalists contributed to the design of the microwave spectrum auction by the Federal Communications Commission (held periodically since 1994), which has raised more federal revenue than has been spent on economic research by all government agencies combined. Experimentalists have developed “designer markets” for research slots on the space shuttle ([Ledyard, Porter, & Wessen, 2000](#)), pollution permits (e.g., [Cason & Plott, 1996](#)), wholesale electricity (e.g., [Smith, Rassenti, & Wilson, 2002](#)), and kidneys ([Roth, Sönmez, & Ünver, 2004](#)), just to name a few. [Plott \(1994\)](#), who is responsible for the term “designer markets,” provides more examples. I believe these are the two main contributions of experimental economics to date: (1) the challenge to the naive rational actor model and the resulting development of theory that incorporates social preferences and bounded rationality; and (2) the study and design of economic institutions.

It is worth mentioning that experimental economists are of course not alone in their emphasis on the importance of institutions. I mention two, both of whom were awarded the Nobel Memorial Prize in Economic Sciences for their work: Elinor Ostrom, who is probably best known for her work on institutions that

---

1. Personal history: my first public act as an academic economist was to discuss an experimental paper presented by Tom Palfrey at the Southern Economic Association meetings in the fall of my third year in graduate school. I commented favorably on the paper (it was published as [Forsythe, Palfrey, & Plott \(1982\)](#) and has become a classic asset market experimental methodology). My first experimental research paper, published 10 years later, was written with Charlie Holt, whom I had met when he was visiting the University of Virginia ([Eckel & Holt, 1989](#)). The one I presented at ESA in 1992 was my first foray into what became the “social preferences” or “fairness” agenda. (It was eventually published as [Eckel & Grossman, 2001](#)). Since then, my research has strayed pretty far from standard economics: at times, I have imagined both Charlies (Holt and Plott) shaking their heads in dismay.

help address public goods and commons problems (e.g., Ostrom, 1998; Ostrom, Gardner, & Walker, 1994), and Douglas North for his work on historical institutions (North, 1990).

Due to the requirement that subjects be paid (discussed in more detail later), experimental economics is more dependent on research funding than most other fields of economics, and the National Science Foundation (NSF) has played a vital role in the growth of the field. In the early years in particular, NSF's support was very important for experimental research and contributed greatly to its growth. The first experimental grant was to Tom Schelling at Harvard in 1960, and Vernon Smith received the second, titled "Behavior in Competitive Markets," in 1962. It was some years before Martin Shubik received the third in 1969. Charlie Plott was first funded in 1972 for a proposal titled "Political Economic Decision Processes," and John Kagel, Ray Battalio, and Robert Basmann at Texas A&M University were funded in 1972 for a proposal that included the first Kagel/Battalio experiments with animals, which later won Senator William Proxmire's "Golden Fleece" award (for an example of this research, see Battalio, Kagel, & MacDonald, 1985).<sup>2</sup> Soon thereafter, in 1974, Dan Newlon became program director for the economics program at NSF. Newlon appreciated experimental research, and under his guidance the review panel has included at least one experimentalist since then. I do not think Newlon realizes how important his support has been for the development of the field.<sup>3</sup>

It was more than 10 years after the founding of ESA that the association finally established its own journal, *Experimental Economics*, and that journey was a rocky one. There was opposition to starting the journal from the members of the ESA, primarily because of the concern that a dedicated journal would provide editors of top economics journals with an excuse to reject experimental papers. Charlie Holt, ESA president from 1991 to 1993 and one of the founding editors of the journal, remembers bringing a motion before the ESA board during his presidency and having it rejected. Finally, in 1997, Holt and Arthur Schram (as co-editors) decided to go ahead with the journal with or without ESA support, but they still hoped the association would go along. Under the leadership of Tom Palfrey, the president at the time, ESA not only adopted the journal but also became fully international, with a structure that included a president, a North American vice president, and a European vice president.<sup>4</sup> I served two terms as vice president (2000–2004), largely because no one remembered to have elections the year my term was up, so it was agreed we would just continue. John Hey as European vice president was in the same situation. In the middle of our double term, Vernon Smith, together with psychologist

---

2. Senator William Proxmire, a democrat from Wisconsin, began issuing the award in 1975 to public officials or projects that he considered to be a waste of public money (personal communication with Dan Newlon, Director of the Economics Program, NSF, October 29, 2006).

3. Personal communication with Charles Plott, October 31, 2006.

4. Personal communication with Charlie Holt, October 28, 2006.

Daniel Kahneman, won the Nobel Prize in economics, the first prize given explicitly for experimental research. We were proud and happy.<sup>5</sup>

## II METHODOLOGY

As is true of any academic field, experimentalists are fussy about their methodology. We have two unbreakable rules: subjects must be paid based on their decisions, and there can be no deception of any kind. The rules are enforced by reviewers and editors who refuse to publish exceptions. This rigid approach may seem a bit extreme to experimentalists in other social sciences, but the rules are there for a reason.

Economists are a tough and skeptical audience, and their shared view of the world is based on economic theory. As theorists [Milgrom and Roberts \(1987\)](#) note, “No mere fact ever was a match in economics for a consistent theory” (p. 185). Presenting an experimental paper in a mainstream economics department used to be hazardous duty. Both market experiments and games came under fire. The criticism of market experiments was always, “How can you learn anything about how markets work from what undergraduates do in a lab for ten bucks?” For games such as public goods or bargaining games in which results nearly always deviated from the game theory, the question was, “Would this hold up with high stakes?” followed by “How can you learn anything about bargaining from what undergraduates do in a lab for ten bucks?” To add credibility to my results, I often began my talks by conducting the game in my experiment with the audience. In one such pretalk game, an ultimatum game (see later discussion), a well-known theorist stood up and shouted to everyone in the room that he would reject anything below 50%, illustrating by example the result I was about to present: that small offers are rejected in this game, in contradiction to the standard game theoretic analysis. All this is by way of making a case that our methods had to be strong enough to convince the rest of the profession that they should pay attention to experimental research. Naturally, economists require substantial monetary stakes in experimental games. This practice continues, even though it has been demonstrated that stakes are not always important. ([Camerer and Hogarth \(1999\)](#) provide a survey of when and how much stakes affect behavior in experiments.)

The prohibition against deception arose as part of the same project to convince the profession that experiments constituted serious and important research. To those outside the field, the term “experimental research” evokes extreme examples of deception, such as the Tuskegee medical experiments or the psychology study conducted by Stanley [Milgram \(1963\)](#), in which subjects

---

5. Not everyone knows that Vernon Smith is a dedicated (and wonderful) country-Western dancer. Charlie Holt recalls being in a cowboy bar one night, with Vernon, as usual, dancing with all the ladies. One woman Charlie danced with said to him, “Isn’t he that guy who’s going to win the Nobel Peace Prize?”

were led to believe they were seriously harming their counterparts in the experiment. The popular perception of experiments in the other social sciences, especially psychology, is that deception is commonplace, indeed more the rule than the exception.

From an economist's perspective, the most serious deception involves payment, such as when an experimenter says he will pay subjects their earnings in a negotiation and then in the end pays everyone a fixed participation fee instead. Particular care is taken with regard to the design of the incentives in an experiment because this is the way the details of the theory are implemented in the lab (as explained later). Experimentalists feared that subjects would not believe they were playing the game that was set up in the lab and that they (the experimenters) would lose control over the subjects' motivation. Control over motivation through payoffs is probably the most critical element of an economics-style experiment. Thus, the prohibition against deception was born. In a sense, a subject pool that believes it will not be deceived is a public good, and even a little deception in one experiment may contaminate the pool and damage the public good. In recent years, there has been a heated debate about the value of deception and its impact on subsequent research; see [Sell \(2008\)](#) for an introduction. A focus article in *Behavior and Brain Sciences* ([Hertwig & Ortmann, 2001](#)) and the associated commentaries further investigate the issues. Finally, [Jamison, Karlan, and Schechter \(2008\)](#) test the notion that deception will contaminate the subject pool. They examine the effect of deception on the subsequent behavior of subjects and find that deceived subjects are less likely to return for future experiments, and those who do return exhibit behavioral differences. So the economists may be onto something here.

To test an economic theory in the lab usually requires a kind of translation of the theory into a set of instructions and protocols—and payments. For example, suppose you want to test supply and demand, the bedrock of Economics 101. When we teach supply and demand, we talk about where demand comes from (buyers' valuations of the good, taking income and all other goods' prices as given) and where supply comes from (sellers' costs), and we ask students to imagine them coming together. Equilibrium in the market is something we explain to students: we ask them to reason out what would happen when demanders with different values and sellers with different costs come together and conclude that the lower-cost sellers and higher-value buyers will end up with the good, all buying or selling at the equilibrium price—the one at which the quantity sellers want to sell agrees with the quantity the buyers want to buy (i.e., where the curves cross—perhaps with the aid of a Walrasian auctioneer). But to test supply and demand in the lab requires that buyers have real values and sellers have real costs and that there be some real mechanism for them to come together and trade. Enter two of the experimentalist's building blocks: *induced values* and *institutions*.

Induced value theory was developed by [Smith \(1982\)](#) to drive home the point that as long as subjects care about money and do not care too much about

other things, we can use the structure of payoffs to induce values—in this case, buyers’ values and sellers’ costs—so as to be able to test theory contingent on those values. The institution is the set of rules that determine who can do what in the experiment. This includes rules about the message space (what information can flow—bids and offers in this case) and who can send what messages (buyers can bid, and sellers can ask), as well as rules about prices and trades (in this case, if a buyer and seller agree to trade at a price, the trade occurs). Induced values and institutions go a long way toward accomplishing the translation of a theory into the lab.<sup>6</sup>

In the case of demand and supply, the experimenter has to decide how many buyers and sellers there are, how many “units” each can buy or sell, and their values or costs. Buyers’ values are induced by paying subjects the difference in cash between their induced value (determined by the experimenter) and the price at which they purchase a unit (determined in the market); sellers earn a more intuitive profit of selling price less (induced) unit cost. How to implement the institution—the rules by which information is exchanged and trade takes place—is not obvious, however. In the supply and demand world of Economics 101, there are no institutions: demanders and suppliers come together as if by magic. But there is no magic in the lab because that would be deception. Instead, the experimenter must design and implement an institution. Sometimes the institution is part of the theory, but very often it is not, as in the case of supply and demand. This brings me to the first canonical game: the double auction.

## A Canonical Games I: The Double Auction

In 1962, Vernon Smith reported the results of a classroom experiment testing supply and demand. According to published histories (e.g., [Eckel, 2004](#)), while a graduate student at Harvard, Vernon participated in a classroom experiment conducted by Edward Chamberlin. [Chamberlin \(1948\)](#) conducted bilateral trading experiments with his graduate students at Harvard to “prove” the failure of the competitive model. He assigned values to buyers and costs to sellers, and the trading “institution” consisted of having subjects wander around the room and negotiate individual trades. The transactions were then summarized: price dispersion disproved the competitive model. He concluded that “economists may have been led unconsciously to share their unique knowledge of the equilibrium point with their theoretical creatures, the buyers and sellers, who, of course, in real life have no knowledge of it whatever” (p. 102). That is, only theoretical buyers and sellers have sense enough to behave in accordance with equilibrium theory.

Vernon Smith, then in his first appointment at Purdue, decided to modify the institution slightly to make it more like stock, bond, and commodities markets

---

6. [Plott \(1979\)](#) refers to the “fundamental equation” of experimental economics, by which preferences and institutions produce outcomes data.

by having buyers and sellers call out prices and providing a pit boss to coordinate things. A second innovation was to repeat the market, in the sense of the movie *Groundhog Day* (Ramis, 1993); every trading “day” started fresh, with buyers and sellers holding the same set of values and costs as the previous day. In his experiment, the supply and demand model did a great job of predicting the outcome in the market. When Vernon tells about this experiment, he maintains he was astonished at the rapid convergence to competitive equilibrium that he observed, and he says he had to replicate it a few times, with different shaped demand and supply arrangements, to convince himself it was correct. Experimentalists now know that the institution he implemented, the oral double auction, is about the most powerful institution there could be for inducing convergence to competitive equilibrium. It works like a charm, with groups as small as three buyers and sellers; even with low or hypothetical stakes, it reliably converges to equilibrium. Every experimentalist I know conducts this experiment in classrooms in almost every class they teach. It is an unforgettable experience for students. They come away believing that markets (can) work, and that the theory of supply and demand has teeth (or jaws, as Plott would say).<sup>7</sup>

The second most popular market institution is probably the posted offer market. Demand and supply are induced as before. In this institution, however, prices are posted by one side of the market (typically sellers), and then buyers are chosen randomly to “go shopping.” Because most of the action is on the seller side of the market, buyers are often simulated in posted offer markets. A comparison of these institutions has produced a remarkable result. In the double auction, market structure has little effect on price. Even a monopoly seems unable fully to exercise market power in this setting. However, in a posted offer market, a small number of sellers can much more easily enforce a price above competitive equilibrium. This very well-established result is something that is not taught in Economics 101. Economists teach that market power is important, but we should also teach that *institutions matter*—the rules of trade have a lot to do with whether and how much market power can be exercised.

The double auction has been used in hundreds of experiments, and the results comparing double auction and posted offer experiments have led to a large body of research comparing theoretically equivalent institutions. For example, in theory, the English increasing-price auction and the Vickrey second-price auction are equivalent: both should lead people to bid up to their values. However, in practice this is not the case. The latter leads to higher prices, as agents systematically overbid their values. Again and again, we learn that institutions matter. It is this close attention to institutions that gave experimentalists

---

7. Plott has a wonderful presentation in which he illustrates the dynamics of price movement using an animated, visual representation of the book (outstanding bids and offers) in an auction market (available online at <http://eeps.caltech.edu/anim.html>). A discussion can be found in Bossaerts and Plott (2008), and recent applications are presented in Plott, Roy, and Tong (2012).

the tools to design markets, as mentioned in the introduction. This may prove to be the greatest contribution of experimental economics, and it is still very much work in progress.<sup>8</sup>

Recent years have seen a remarkable expansion in market experiments, focused largely on institutions and how they interact with imperfect information. Several surveys focus on developments in asset markets (Noussair & Tucker, 2013; Palan, 2013), auctions (Kagel & Levin, 2009; Kwasnica & Sherstyuk, 2013), and prediction markets (Deck & Porter, 2013). Market design, which combines game theory and experimental research to design new markets, has also seen tremendous growth, reflected in and spurred by the decision to award the 2012 Nobel Memorial Prize in Economic Science to Alvin Roth and Lloyd Shapley (Nobelprize.org, 2013). Here, lab experiments provide a kind of “wind tunnel” for testing behavior in new institutions, and field experiments allow for further refinement of the institutional mechanisms. New allocation markets include those for matching physician interns (Niederle & Roth, 2005; Roth, 1991), kidneys (Roth et al., 2004; Sönmez & Ünver, 2010), school choice (Chen & Kesten, 2012; Featherstone & Niederle, 2011), and the job market for new economists (Coles et al., 2010). Clearly, this is a fruitful and growing area with enormous practical potential.

## B Canonical Games II: The Public Goods Game

The first public goods experiment was conducted by Peter Bohm (1972). His objective was to elicit willingness to pay for a television show, and he cautiously concluded that the public goods problem—the incentive to conceal one’s valuation for public goods—was overstated. The experiment had flaws, but it was the first to find that subjects would contribute to a public good. The canonical implementation of the public goods situation in economics came later: it is the linear voluntary contribution mechanism, first studied by Isaac, Walker, and Thomas (1984). The game is essentially a multiperson continuous prisoner’s dilemma, a game that is well-known to every social scientist. Ledyard (1995) provides the definitive survey of early research, and Vesterlund (forthcoming) provides a survey focused on several key recent developments in public goods research.

Several features of Isaac et al. (1984) have become standard practice in public goods research. In this game, subjects are brought into the lab and formed into groups, typically of 4 participants, although experiments with group sizes of 2–10 are common, and larger groups have also been studied. Each subject is given an endowment, usually of “tokens,” that can be exchanged for dollars at a predetermined exchange rate. The endowment can be invested in one of two accounts—the individual account or the group account. The individual account pays off one token per token invested; the group account pays each person in

---

8. For surveys of experimental research on market institutions, see Davis and Holt (1993, Chapters 3 and 4) and Kagel and Roth (1994, Chapters 5–7).

the experiment a fraction of a token—less than 1 but greater than  $1/n$ , where  $n$  is the group size ( $n=3$  in this study). This fraction is called the marginal per capita return (MPCR). The game is usually repeated, and 10 rounds is the typical length, although longer and shorter versions have been conducted.

This game mimics the incentive structure of a public good: consumption is nonexcludable and nonrival. Each token invested in the group account produces one unit of public good worth MPCR to each person. Because most of the benefit of the public good accrues to others, there will be underprovision of the good relative to the social optimum, which is for all tokens to be contributed to the group account. The tension of course is the same as in the prisoner's dilemma game. Because  $MPCR > 1/n$ , cooperation (investing all one's tokens in the group account) maximizes efficiency and leads to higher payoffs than if no one invests. However, because  $MPCR < 1$ , each person has an incentive to free-ride off others' contributions and invest all his or her own tokens in the individual account.

The Nash equilibrium of the game (assuming payoff-maximizing agents) is for all players to contribute zero to the group account, but the results of the experiments differ from this prediction and are very stable: contributions average 40–60% of the endowment in early rounds and deteriorate close to zero by the 10th round. There is also considerable heterogeneity across individuals and for a given individual over time. This is the first game to show so much variation across individuals. In the market games, the only variation in play comes from the differences in induced values provided by the payoff structure of the experiment. Here, there is clear variation in what people do.

The provocative results of early public goods game experiments were followed by the publication of a large number of clever variations on the game. Many studies were designed to determine why subjects behave in this way, contributing to the production of public goods in contradiction to the rational actor model. Although some may term it irrational, it is difficult to sustain that opinion when subjects succeed in extracting considerably higher earnings from the experimenter than they would from uniform free-riding. Hypotheses include mistakes, confusion, altruism, reciprocity, social norms, and conditional cooperation.

Economists have responded to experimental data showing cooperative behavior by developing new formal theory. Three particularly important models alter the utility functions of players, explicitly modeling the ways in which individual subjective payoffs may diverge from simple monetary payoffs. Two such models focus on the differences between own and others' payoffs (Bolton & Ockenfels, 2000; Fehr & Schmidt, 1999). Both model agents as “inequality averse”—that is, they lose utility when own payoffs are different from others. These models can generate predictions consistent with cooperative play in the public goods game. Charness and Rabin (2002) also add a preference for efficiency to the utility function. All of these models fit the data better than the naive, monetary payoff-maximizing rational actor model. During approximately

the past decade, papers dealing with such “social preferences” have flourished. This work is surveyed in [Cooper and Kagel \(forthcoming\)](#).

Nevertheless, this game was the first to illustrate the obvious holes in the rational actor model, narrowly drawn. This is not to diminish the value of rational choice analysis. As [Schotter \(2006\)](#) argues, the main value of building models based on the assumption of rational actors is that it allows us to separate logical “wheat” from intuitive “chaff” by carefully proving theorems about the implications of assumptions. Schotter states, “[P]eople get confused by the rational choice methodology when they believe that the results of the theorems proven in this fashion are correct predictors of human behavior. Mature thinkers understand that this is not true” (p. 500). When I claim this game’s results revealed holes in the rational actor model, I do not mean to criticize the modeling exercise; like Schotter, I think we learn a lot from theory that is “strong and wrong.”

However, many economists do seem to swallow their theory whole, if not as a predictor of human behavior, then as a normative description of how behavior should be conducted. I have so often seen an economics professor teach innocent students in principles of economics that it is “irrational” not to free-ride in a prisoner’s or social dilemma situation. Indeed, a fictional account of just such an event appears in Jane Smiley’s novel *Moo* (1995). Although baffling to Smiley’s (female) character, to me this exercise is somewhat irresponsible and a missed opportunity. Instead, we should be teaching students that situations such as this signal caution: awareness of the free-rider problem should be used to solve it, not worsen it by encouraging free-riding. Some students free-ride by nature: in experiments, approximately 25% are natural rational actors. However, most pursue different strategies in these games. A few are principled cooperators and always contribute everything. More than half of subjects exhibit what looks like conditional cooperation: they cooperate when others do so, and they respond to free-riding with more of the same. Indeed, these conditional strategies earn the most money for their practitioners when played across many different groups of players. When prompted to choose a multiperiod strategy for playing the public goods game, [Keser \(2000\)](#) found that the most successful strategies were contingent, depending on the past actions of others in the group.

Considerable progress also has been made in the study of institutions to mitigate public goods situations, again using as their foundation the canonical public goods game. Notable is the recent surge in research on the role of punishment in encouraging or enforcing cooperation. Although punishment can be viewed as a kind of public good itself—in the sense that the punisher incurs a cost that produces a benefit for all others—experiments that allow punishment show that subjects engage in punishment behavior, and this leads to higher levels of cooperation. The seminal papers in this line of research are [Yamagishi \(1986\)](#) and [Fehr and Gächter \(2000\)](#). See also the survey by [Chaudhuri \(2011\)](#).

## C Canonical Games III: The Ultimatum Game

When three German economists conducted the first ultimatum bargaining game experiment ([Güth, Schmittberger, & Schwarze, 1982](#)), their purpose was to strip bargaining down to its essentials by creating the simplest possible bargaining situation. In this game, there are two players, usually called the proposer and the responder. An amount of money is made available to the pair. The proposer's task is to determine a division of the money. The responder is then presented with this proposal, and his or her task is to accept or reject the proposal. If accepted, the money is divided as proposed; if rejected, both players receive zero earnings, and the money reverts to the experimenter. The subgame perfect equilibrium of the game is for the payoff-maximizing responder to accept any positive offer and for the proposer, knowing that, to offer the smallest positive increment available. For example, in a \$10 ultimatum game with decisions restricted to dollar amounts, the equilibrium would involve payoffs of \$9 to the proposer and \$1 to the responder.

This game might seem overly artificial, with little relevance to bargaining behavior in the field. On its own, this is a valid criticism. However, the ultimatum game plays an important part in theory: this game makes up the final stage of many multistage games, and it is relied on by much of principal–agent theory as the final stage of a transaction (i.e., the principal pays the agent his or her reservation wage, and the agent accepts). Thus, behavior in the ultimatum game has implications for a broad set of important economic problems involving worker motivation, contracting, and the notion of a fair price.

[Güth et al. \(1982\)](#) found that very few proposers proposed such an unequal division, and that such divisions were routinely rejected. The modal offer was 50% of the pie for inexperienced subjects and 40% for experienced. In a second experiment, subjects played both sides of the game, stating an offer as proposer and a minimum acceptable offer as responder. The results were essentially unchanged. When Güth et al.'s results were published, the reaction of many economists was to assume they had done the experiment wrong. It was treated as a mere anomaly, not a true finding. There followed dozens of studies attempting to do it correctly ([Camerer & Thaler, 1995](#); [Thaler, 1988](#)). The real issue was not the behavior of proposers, because they might be maximizing their expected payoffs conditional on expectations about the pattern of rejections. But how can one explain the behavior of responders? An income-maximizing responder would accept any positive offer, and yet offers of 30% or less of the pie were rejected.

Were the subjects confused? One study ([Kahneman, Knetsch, & Thaler, 1986](#)) explained the game, gave subjects a quiz, and kicked out all the ones who did not give the right answer. Offers and minimum acceptable offers were approximately the same. Later studies attempted to “teach” second movers to accept low offers by repeating the game. [Ochs and Roth \(1989\)](#) repeated the game 10 times, randomly rematching subjects each round. The results of these

experiments were very similar to those of the one-shot games. (See also the studies cited in [Kagel and Roth, 1995](#), Chapter 4.)

Methodological detour: random rematching is one of two main methodological innovations resulting from the ultimatum game agenda. It is a technique that experimental economists use to give subjects an opportunity to learn about a game without turning it into a repeated game. Repeated games are problematic because they generate so many possible equilibria. Each subject is matched with a different counterpart each round to replay the game. In the most careful such designs, subjects never meet a counterpart twice, nor a counterpart who has played with one of their counterparts. This is tricky and probably in vain because some evidence suggests that subjects respond approximately the same to information whether it comes from a one-time or repeated counterpart ([Ashley, Ball, & Eckel, 2010](#)).

Was this anomaly the result of low stakes? You will recall that this is a very common seminar challenge. [Hoffman, McCabe, and Smith \(1995\)](#), weary of the challenge, spent \$5,000 to find out. They played the game with \$100 stakes. The proportional distribution of offers was not significantly different from the \$10 version of the game, and more surprisingly, three of four offers of \$10 and two of three offers of \$30 were rejected. [Cameron \(1999\)](#) conducted very high stakes experiments in Indonesia and found essentially the same result. [Slonim and Roth \(1998\)](#) combine learning and stakes. Subjects play low- and high-stakes games, with payoffs between 60 and 1,500 Slovak crowns,<sup>9</sup> repeated for 10 rounds. In their study, the rejection rate falls slightly with higher stakes and with repetition, and offers tend to converge toward approximately 40% of the pie.

Were the subjects punishing unfair behavior? [Kahneman et al. \(1986\)](#) report another study that focused on punishment. In the first stage of the game, subjects could divide a \$20 pie only one of two ways: by keeping \$10 and giving \$10 to their counterpart or by keeping \$18 and giving \$2. The counterpart was passive and did not have the opportunity to reject the offer. Seventy-six percent split it equally. (I believe this is the first dictator game, in which the respondent is a passive recipient of the proposer's offer, although they did not refer to it as such.) In the second stage of the game, subjects were matched with two recipients and could divide \$10 (evenly) with someone who had previously kept \$10 or divide \$12 with someone who had previously kept \$18. In this game, 74% chose the smaller pie, sacrificing \$1 in payoffs to reward the fair counterpart. Clearly, fairness plays a role, and subjects are willing to sacrifice small amounts of money to punish unfair behavior. My favorite illustration of the importance of fairness comes from Sally [Blount \(1995\)](#). Her study compares offers randomly generated by computer with those generated by the proposer. When the

---

9. The exchange rate at the time was approximately 30 Crowns per US dollar, so the stakes were substantial.

computer generates the offer, unfair offers are readily accepted by responders. However, the same offer generated by the proposer is rejected as usual.

Did proposers fear rejection, or were they being deliberately fair? [Forsythe, Horowitz, Savin, and Sefton \(1994\)](#) explicitly compare the ultimatum game with a dictator game, in which the proposer can make the same set of offers as in the ultimatum game, but the responder must passively accept. In the dictator game, there is no risk of rejection, so fairness is the only possible motivation for positive offers. They find that with hypothetical decisions, there is little difference between the two games, but when stakes are positive (\$5), the dictator game results in substantially more selfish behavior, with approximately 40% of subjects keeping the full endowment. Increasing the stakes to \$10 has little additional effect.<sup>10</sup>

Is anonymity a factor? In game theory, agents are essentially anonymous. They have no individual or social identities, in the way psychologists or sociologists think of identity, and their counterparts do not either. An implication of this is that agents do not care if they are observed, and they do not care about the identity of the person with whom they are playing. However, in the experiments we conduct, subjects are not anonymous. They know who else is in the room, and they know the experimenter is going to see what they do. One hypothesis, then, is that subjects are fair because they are observed.

[Hoffman, McCabe, Shachat, and Smith \(1994\)](#) introduce the second methodological innovation resulting from the ultimatum game agenda: the double-blind or double-anonymous protocol. In this protocol, subjects know that their actions are not observed, and they are anonymous to each other. The experiment is designed in such a way that the experimenter ends up with a set of proposer decisions that cannot be attributed to any particular person, and the responders cannot determine who their offers came from. Their experiment is a dictator game, and the result was a notable increase in selfish play, with approximately 70% of subjects keeping the full endowment.

A notable property of ultimatum game data, like data from the public goods game, is a great deal of heterogeneity in behavior. This game and its descendant, the dictator game, both show a great deal of variation in individual play. In the early 1990s, researchers began to realize that this property made them useful as instruments for measuring preferences. [Eckel and Grossman \(2001\)](#) and [Solnick \(2001\)](#) investigated differences in the behavior of women and men in the ultimatum game, and [Eckel and Grossman \(1998\)](#) did the same for the dictator game. The ultimatum game was adopted as a tool for measuring social norms in a study comparing behavior across 15 small societies ([Henrich et al., 2004](#)). The use of experiments as measures of preferences has developed on its own path since. I return to this discussion later.

---

10. The dictator game is arguably another canonical game. We return briefly to it in the concluding section. A survey of ultimatum and dictator games is available in [Camerer \(2003\)](#), and a meta-analysis of dictator game experiments can be found in [Engel \(2011\)](#). See also [Güth \(1995\)](#).

## D Canonical Games IV: The Trust Game

The final game I highlight is the “investment game,” first presented by [Berg, Dickhaut, and McCabe \(1995\)](#), now referred to primarily as the “trust game.” In this game, two players are endowed with \$10. The first mover can send any amount between \$0 and \$10 to the second mover. On the way, any amount sent is tripled so that if the first mover sends \$1, it becomes \$3, and \$10 becomes \$30. The second mover can then decide to return any amount between \$0 and the full amount he or she received. (The returned amount is not tripled.) This game is interesting because of its sequential nature. Like the public goods game, there are gains to cooperation, but to achieve those gains the first mover must first trust by putting his or her payoffs in the hands of the second mover, with no promise of return. The amount sent by the first mover is “trust,” and the amount returned is “reciprocity.”

Similar to the previous games, the Nash equilibrium of the game is for the second mover to keep any money sent to him or her and for the first mover, knowing that, to send zero. But this is not what happens in the game. [Berg et al. \(1995\)](#) implemented a double-anonymous protocol, similar to the one invented by [Hoffman et al. \(1994\)](#), which should enhance self-interested behavior. To their surprise, subjects sent on average approximately half of their \$10 endowments, and trust nearly paid, with slightly less than \$5 returned on average.

This game was soon adopted as a way to measure trust at the individual level. Because the notion of generalized trust had recently been popularized by [Fukuyama \(1995\)](#), and levels of trust measured throughout the world using several questions from the World Values Survey, it seemed natural to examine this game along with the survey questions typically used to measure it. [Glaeser, Laibson, Scheinkman, and Souter \(2000\)](#) conducted a variation on this experiment with Harvard undergraduates, and they found a very high level of trust and reciprocity, a great deal of individual variation, and essentially no correlation with the survey questions. However, behavior in the game was correlated with trusting and trustworthy behavior in several domains.

This game has been replicated dozens of times with different populations throughout the world, and as with the ultimatum game, many of the studies focus on the question of why people behave as they do in the game. However, the tone of these studies is very different from that of the early public goods and ultimatum games. Economists are not asking, “Why are people behaving so stupidly?” Instead, attention is paid to determining the mechanisms that affect behavior in the game and how behavior is related to underlying preferences such as altruism, as measured by the dictator game (e.g., [Burns, 2006; Capra, Lanier, & Meer, 2008; Cox, 2004](#)), or risk aversion, as measured by behavior in risky-choice tasks ([Eckel & Wilson, 2004; Houser, Schunk, & Winter, 2010; Schechter, 2007](#)). A recent meta-analysis examines determinants of trust and reciprocity ([Johnson & Mislin, 2011](#)).

### III DISCUSSION AND CONCLUSION

Experimental economics developed as a way to test theory under controlled conditions, especially theories that could not be easily tested with available field data. Although experimental research initially was met with skepticism, over time it has become accepted as a legitimate, indeed important, element of the economist's toolkit. With the awarding of the Nobel Prize in economics to Vernon Smith and Daniel Kahneman in 2003, the field officially came of age; both Nobel lectures were published in the *American Economic Review* (Kahneman, 2003; Smith, 2003). Subsequent prizes were awarded to other prominent experimentalists, including Elinor Ostrom in 2009 for her work on governing the commons (Ostrom, 2010) and Alvin Roth in 2012 for market design.<sup>11</sup>

My particular interest in experimental research is twofold, and both interests are different from the primary impetus for experimental research—testing theory. The first is the use of experimental games as tools for measuring preferences; the second is the study of heterogeneity in behavior.

Measuring preferences has historically been in the domain of survey research. However, experiments may be superior to surveys. Self-reported survey measures of preferences have shortcomings. From an economist's point of view, the trouble is that a person has no particular incentive to reveal his or her preferences truthfully and may have an incentive to misrepresent them. Suppose, for example, that you are faced with a survey question that asks you to rate your altruism on a scale of 1 to 5. If you are in an economics class, you may want to show your professor how rational you are by choosing 1 as an answer. Or if you think your answer will be observed by an experimenter whose primary interest is in showing how nice people are, you might be tempted to give her the answer she wants and mark a 5. There is no cost to misrepresenting your true preferences. In addition to impressing others, you may choose a decision that is consistent with how you wish you behaved, consciously or unconsciously using the survey answer to affirm your self-image.

Economists' view of preferences as represented by a utility function with certain properties suggests an alternative measurement strategy. Altruism can be measured as the trade-off between own and others' payoffs. Consider the dictator game, in which the recipient is an anonymous person, or better yet a reputable charity, as an alternative measure. In order to show your altruism, you must forego income. Showing altruism is not free: you must actually behave altruistically.

This approach may yield a measure with a higher degree of external validity, although the jury is still out on that. Very few studies compare survey and experimental measures, but those that do tend to find low levels of correlation

---

11. I also mention Reinhard Selten, who shared the prize in 1994 with John Nash and John Harsanyi for game theory but also was a pioneer experimental economist.

between survey and experimental measures, and some find greater predictive power for the experiments (e.g., [Glaeser et al., 2000](#)).<sup>12</sup>

Using experiments to study individual differences is another focus of my own work. Heterogeneity arises from two sources. First, people behave differently from each other. For example, women might be more altruistic and tall people less cooperative on average than others. Second, a given person may behave differently depending on who she interacts with. Experiments have an advantage in studying how people interact. Consider racial discrimination. In an experiment, we can place subjects into a fairly natural situation in which they are matched with another person and have to make a decision. This experiment can be set up without mentioning race or discrimination, and decisions can be made without knowing what the study is about. In a study using the trust game, for example, Rick Wilson and I found evidence that the color of a counterpart's skin affects expectations, trust, and reciprocity ([Eckel, 2007](#)). In this protocol, we match people over the Internet to play a trust game. We reveal skin shade (and gender, attractiveness, etc.) by showing each player his or her counterpart's photograph. The game happens in real-time (no deception, remember). In this way, we can reveal critical information without compromising anonymity or introducing the possibility of postgame interaction. Because discrimination is not an obvious element of the experiment, subjects may reveal latent or unconscious discrimination by their behavior. We also found evidence that beauty affects trust and reciprocity, primarily through its impact on expectations about a counterpart's behavior ([Wilson & Eckel, 2006](#)). Indeed, subjects value the information contained in a counterpart's photo and are willing to pay to see the photo ([Eckel & Petrie, 2011](#)).

An important application of this research is the study of gender differences. This is also an area that has seen considerable growth during approximately the past decade. [Croson and Gneezy \(2009\)](#) provide a survey of several major lines of research, and they conclude that women are on average more averse to risk, more willing to share in the dictator game, and more trustworthy, but they behave similarly in other settings. An important line of research documents the stark gender differences in response to competition: women tend to avoid competition, and this may significantly affect labor market outcomes ([Niederle & Vesterlund, 2007](#)).

In addition to the measurement issues discussed previously, a list of contributions of experimental economics, or what we think we know, would have to include at least the following.

- Noncooperative game theory is a powerful tool for predicting behavior in competitive environments. Competition seems to enforce self-interest, producing agents who look very much like economic man. In these settings,

---

12. Missing from this chapter is a discussion of individual decision-making under uncertainty. This large area of research is also grappling with the problem of eliciting preferences. A good discussion is provided by [Holt and Laury \(2005\)](#).

which include a wide range of relevant environments, the rational actor model works pretty well at predicting outcomes.

- Institutions matter. That is, the exact rules about what and how information can be exchanged, and how trades take place, can have a very strong impact on economic outcomes. In particular, the double oral auction is a very efficient institution, probably because it is clearly competitive and a great deal of information is exchanged in the bidding process. The posted offer auction is less obviously competitive and less information comes out in the bidding process, making market power more easily exercised in this setting.
- People do not always behave “rationally.” Although some institutions appear to enforce rationality, behavior that is other-regarding—such as altruism, cooperation, and spite—is observed in less “disciplining” institutions. Games in which others’ payoffs are salient, and especially where there are potential gains to cooperation, are especially likely to elicit other-regarding behavior. Public goods and trust games are examples.
- When game theory is predictive, it is sometimes for the wrong reason. Nash assumed that people knew each others’ preferences and tastes (utilities), not merely their payoffs. However, we are more likely to see behavior leading to the Nash equilibrium when people do not know others’ payoffs. More information improves decisions in some settings; but if it removes anonymity, it seems to change the game as subjects weigh others’ payoffs in their decision-making calculus.

## ACKNOWLEDGMENTS

I thank Angela C. M. de Oliveira for input; Charles Holt, Dan Newlon, and Charles Plott for their memories; and Jane Sell and Rick Wilson for guidance.

## REFERENCES

- Ashley, R., Ball, S., & Eckel, C. (2010). Motives for giving: A reanalysis of two classic public goods experiments. *Southern Economic Journal*, 77(1), 15–26.
- Battalio, R. C., Kagel, J. H., & MacDonald, D. N. (1985). Animals’ choices over uncertain outcomes: Some initial experimental results. *American Economic Review*, 75(4), 597–613.
- Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, 10(1), 122–142.
- Blount, S. (1995). When social outcomes aren’t fair: The effects of causal attributions on preferences. *Organizational Behavior and Human Decision Processes*, 63(2), 131–144.
- Bohm, P. (1972). Estimating demand for public goods. *European Economic Review*, 3, 111–130.
- Bolton, G., & Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. *American Economic Review*, 90(1), 166–193.
- Bossaerts, P., & Plott, C. R. (2008). From market jaws to the Newton method: The geometry of how a market can solve systems of equations. *Handbook of Experimental Economics Results*, 1, 22–24.
- Burns, J. (2006). Racial stereotypes, stigma and trust in post-Apartheid South Africa. *Economic Modeling*, 23(5), 805–821.

- Camerer, C. F. (2003). *Behavioral game theory: Experiments in strategic interaction*. New York: Russell Sage Foundation/Princeton University Press.
- Camerer, C. F., & Hogarth, R. M. (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty*, 19(1–3), 7–42.
- Camerer, C. F., & Thaler, R. H. (1995). Ultimatums, dictators, and manners. *Journal of Economic Perspectives*, 9(2), 209–219.
- Cameron, L. A. (1999). Raising the stakes in the ultimatum game: Experimental evidence from Indonesia. *Economic Inquiry*, 37, 47–59.
- Capra, C. M., Lanier, K., & Meer, S. (2008). *Attitudinal and Behavioral Measures of Trust: A New Comparison*. Emory University. Retrieved from: [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1091539](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1091539).
- Cason, T. N., & Plott, C. R. (1996). EPA's new emissions trading mechanism: A laboratory evaluation. *Journal of Environmental Economics and Management*, 30(2), 133–160.
- Chamberlin, E. H. (1948). An experimental imperfect market. *Journal of Political Economy*, 56, 95–108.
- Charness, G., & Rabin, M. (2002). Understanding social preferences with simple tests. *Quarterly Journal of Economics*, 117, 817–869.
- Chaudhuri, A. (2011). Sustaining cooperation in laboratory public goods experiments: A selective survey of the literature. *Experimental Economics*, 14(1), 47–83.
- Chen, Y., & Kesten, O. (2012). From Boston to Shanghai to deferred acceptance: Theory and experiments on a family of school choice mechanisms. In *Auctions, Market Mechanisms, and Their Applications*. New York: Springer.(pp. 58–59). .
- Coles, P., Cawley, J., Levine, P. B., Niederle, M., Roth, A. E., & Siegfried, J. J. (2010). The job market for new economists: A market design perspective. *Journal of Economic Perspectives*, 24(4), 187–206.
- Cooper, D. J., & Kagel, J. H. (forthcoming). Other regarding preferences: A selective survey of experimental results. In J. H. Kagel & A. E. Roth (Eds.), *Handbook of experimental economics* (Vol. 2). Princeton, NJ: Princeton University Press.
- Cox, J. C. (2004). How to identify trust and reciprocity. *Games and Economic Behavior*, 46, 260–281.
- Croson, R., & Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, 47(2), 448–472.
- Davis, D. D., & Holt, C. A. (1993). *Experimental economics*. Princeton, NJ: Princeton University Press.
- Deck, C., & Porter, D. (2013). Prediction markets in the laboratory. *Journal of Economic Surveys*, 27(3), 589–603.
- Eckel, C. C. (2004). Vernon Smith, Nobel Laureate: Economics as a laboratory science. *Journal of Socio-Economics*, 33, 15–28.
- Eckel, C. C. (2007). People playing games: The human face of experimental economics [proceedings paper]. *Southern Economic Journal*, 73(4), 841–857.
- Eckel, C. C., & Grossman, P. J. (1998). Are women less selfish than men? Evidence from dictator games. *Economic Journal*, 108(448), 726–735.
- Eckel, C. C., & Grossman, P. J. (2001). Chivalry and solidarity in ultimatum games. *Economic Inquiry*, 39(2), 171–188.
- Eckel, C. C., & Holt, C. A. (1989). Strategic voting in agenda-controlled committee experiments. *American Economic Review*, 79(4), 763–773.
- Eckel, C. C., & Petrie, R. (2011). Face value. *American Economic Review*, 101(4), 1497–1513.

- Eckel, C. C., & Wilson, R. K. (2004). Is trust a risky decision? *Journal of Economic Behavior and Organization*, 55(4), 447–465.
- Engel, C. (2011). Dictator games: A meta study. *Experimental Economics*, 14(4), 583–610.
- Featherstone, C., & Niederle, M. (2011). *School choice mechanisms under incomplete information: An experimental investigation*. Unpublished manuscript.
- Fehr, E., & Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, 90(4), 980–994.
- Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, 114(3), 817–868.
- Forsythe, R., Horowitz, J., Savin, N., & Sefton, M. (1994). Fairness in simple bargaining experiments. *Games and Economic Behavior*, 6, 347–369.
- Forsythe, R., Palfrey, T. R., & Plott, C. R. (1982). Asset valuation in an experimental market. *Econometrica*, 50, 537–568.
- Fukuyama, F. (1995). *Trust: The social virtues and the creation of prosperity*. Glencoe, IL: Free Press.
- Glaeser, E. L., Laibson, D. I., Scheinkman, J. A., & Soutter, C. L. (2000). Measuring trust. *Quarterly Journal of Economics*, 115(3), 811–846.
- Güth, W. (1995). On ultimatum bargaining experiments: A personal review. *Journal of Economic Behavior and Organization*, 27(3), 329–344.
- Güth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior and Organization*, 11, 417–449.
- Henrich, J., Boyd, R., Bowles, S., Camerer, C. F., Fehr, E., & Gintis, H. (Eds.). (2004). *Foundations of human sociality: Economic experiments and ethnographic evidence from fifteen small-scale societies*. New York: Oxford University Press.
- Hertwig, R., & Ortmann, A. (2001). Experimental practices in economics: A methodological challenge for psychologists? *Behavioral and Brain Sciences*, 24(03), 383–403.
- Hoffman, E., McCabe, K., Shachat, K., & Smith, V. (1994). Preferences, property rights and anonymity in bargaining games. *Games and Economic Behavior*, 7(3), 346–380.
- Hoffman, E., McCabe, K., & Smith, V. (1995). On expectations and the monetary stakes in ultimatum games. *International Journal of Game Theory*, 7, 289–302.
- Holt, C. A., & Laury, S. K. (2005). Risk aversion and incentive effects: New data without order effects. *American Economic Review*, 95(3), 902–904.
- Houser, D., Schunk, D., & Winter, J. (2010). Distinguishing trust from risk: An anatomy of the investment game. *Journal of Economic Behavior & Organization*, 74(1), 72–81.
- Isaac, R. M., Walker, J. M., & Thomas, S. H. (1984). Divergent evidence on free riding: An experimental examination of possible explanations. *Public Choice*, 43, 113–149.
- Jamison, J., Karlan, D., & Schechter, L. (2008). To deceive or not to deceive: The effect of deception on behavior in future laboratory experiments. *Journal of Economic Behavior & Organization*, 68(3), 477–488.
- Johnson, N. D., & Mislin, A. A. (2011). Trust games: A meta-analysis. *Journal of Economic Psychology*, 32(5), 865–889.
- Kagel, J. H., & Levin, D. (2009). *Common value auctions and the winner's curse*. Princeton, NJ: Princeton University Press.
- Kagel, J. H., & Roth, A. E. (1994). *Handbook of experimental economics*. Princeton, NJ: Princeton University Press.
- Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *American Economic Review*, 93, 1449–1475.

- Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1986). Fairness and the assumptions of economics. *Journal of Business*, 59(4, Part 2), S285–S300.
- Keser, C. (2000). *Strategically planned behavior in public goods experiments*. Montreal: CIRANO.
- Kwasnica, A. M., & Sherstyuk, K. (2013). Multiunit auctions. *Journal of Economic Surveys*, 27(3), 461–490.
- Ledyard, J. O. (1995). Public goods: A survey of experimental research. In J. H. Kagel & A. E. Roth (Eds.), *The handbook of experimental economics*. Princeton, NJ: Princeton University Press.
- Ledyard, J. O., Porter, D., & Wessen, R. (2000). A market-based mechanism for allocating space shuttle secondary payload priority. *Experimental Economics*, 2, 173–195.
- Milgram, S. (1963). Behavioral study of obedience. *Journal of Abnormal and Social Psychology*, 67(4), 371.
- Milgrom, P., & Roberts, J. (1987). Informational asymmetries, strategic behavior, and industrial organization. *American Economic Review*, 77, 184–193.
- Niederle, M., & Roth, A. E. (2005). The gastroenterology fellowship market: Should there be a match? *American Economic Review*, 95(2), 372–375.
- Niederle, M., & Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much? *Quarterly Journal of Economics*, 122(3), 1067–1101.
- Nobelprize.org (Producer). (2013, December 22). *Alvin E. Roth—Prize Lecture: The theory and practice of market design*. Retrieved from: [http://www.nobelprize.org/nobel\\_prizes/economic-sciences/laureates/2012/roth-lecture.html](http://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/2012/roth-lecture.html).
- North, D. C. (1990). *Institutions, institutional change, and economic performance*. New York: Cambridge University Press.
- Noussair, C. N., & Tucker, S. (2013). Experimental research on asset pricing. *Journal of Economic Surveys*, 27(3), 554–569.
- Ochs, J., & Roth, A. E. (1989). An experimental study of sequential bargaining. *American Economic Review*, 79, 355–384.
- Ostrom, E. (1998). A behavioral approach to the rational choice theory of collective action. *American Political Science Review*, 92, 1–22.
- Ostrom, E. (2010). Beyond markets and states: Polycentric governance of complex economic systems. *American Economic Review*, 100(3), 641–672.
- Ostrom, E., Gardner, R., & Walker, J. (1994). *Rules, games, and common-pool resources*. Ann Arbor, MI: University of Michigan Press.
- Palan, S. (2013). A review of bubbles and crashes in experimental asset markets. *Journal of Economic Surveys*, 27(3), 570–588.
- Plott, C. R. (1979). The application of laboratory methods to public choice. In C. S. Russell (Ed.), *Collective decision making: Applications from public choice theory* (pp. 137–160). New York: Resources for the Future.
- Plott, C. R. (1994). Designer markets: Laboratory experimental methods in economics. *Economic Theory*, 4(1), 3–10.
- Plott, C. R., Roy, N., & Tong, B. (2012). Marshall and Walras, disequilibrium trades and the dynamics of equilibration in the continuous double auction market. *Journal of Economic Behavior & Organization*, 94, 190–205.
- Ramis, H. (Writer). (1993). *Groundhog day* (T. Albert & H. Ramis, Producers). Culver City, CA: Columbia Pictures.
- Roth, A. E. (1991). A natural experiment in the organization of entry-level labor markets: Regional markets for new physicians and surgeons in the United Kingdom. *American Economic Review*, 81, 415–440.

- Roth, A. E. (1993). On the early history of experimental economics. *Journal of the History of Economic Thought*, 15, 184–209.
- Roth, A. E., Sönmez, T., & Ünver, M. U. (2004). Kidney exchange. *Quarterly Journal of Economics*, 119(2), 457–488.
- Schechter, L. (2007). Traditional trust measurement and the risk confound: An experiment in rural Paraguay. *Journal of Economic Behavior & Organization*, 62(2), 272–292.
- Schotter, A. (2006). Strong and wrong: The use of rational choice theory in experimental economics. *Journal of Theoretical Politics*, 18, 498–511.
- Sell, J. (2008). Introduction to deception debate. *Social Psychology Quarterly*, 71(3), 213.
- Slonom, R., & Roth, A. E. (1998). Learning in high stakes ultimatum games: An experiment in the Slovak republic. *Econometrica*, 66, 569–596.
- Smiley, J. (1995). *Moo*. New York: Knopf.
- Smith, V. L. (1962). An experimental study of competitive market behavior. *Journal of Political Economy*, 70(2), 111–137.
- Smith, V. L. (1982). Microeconomic systems as experimental science. *American Economic Review*, 72, 923–955.
- Smith, V. L. (2003). Constructivist and ecological rationality in economics. *American Economic Review*, 93, 465–508.
- Smith, V. L., Rassenti, S. J., & Wilson, B. (2002). Using experiments to inform the privatization/deregulation movement in electricity. *Cato Journal*, 21(3), 515–544.
- Solnick, S. (2001). Gender differences in the ultimatum game. *Economic Inquiry*, 39(2), 189–200.
- Sönmez, T., & Ünver, M. (2010). Market design for kidney exchange. In N. Vulkan, A. E. Roth, & Z. Neeman (Eds.), *Oxford handbook of market design*. Oxford, UK: Oxford University Press.
- Thaler, R. (1988). Anomalies: The ultimatum game. *Journal of Economic Perspectives*, 2(4), 195–206.
- Vesterlund, L. (Forthcoming.). Voluntary giving to public goods: Moving beyond the linear VCM. In J. H. Kagel & A. E. Roth (Eds.), *The handbook of experimental economics* (Vol. 2): New York: Elsevier.
- Wilson, R. K., & Eckel, C. C. (2006). Judging a book by its cover: Beauty and expectations in the trust game. *Political Research Quarterly*, 59(2), 189–202.
- Yamagishi, T. (1986). The provision of a sanctioning system as a public good. *Journal of Personality and Social Psychology*, 51(1), 110–116.

## Chapter 16

# Solving Coordination Problems Experimentally

Giovanna Devetag

*LUISS Guido Carli, Rome, Italy*

Andreas Ortmann

*University of New South Wales, Sydney, Australia*

### I INTRODUCTION

Coordination problems are pervasive in real life. For example, how do people coordinate on one of many possible meeting places (especially when the mobile network is down temporarily), or how do firms coordinate their investment projects if they cannot communicate about their options? How can bank runs be avoided, or for that matter speculative attacks? More abstractly, how do people coordinate on one of many possible outcomes? Does it help if the possible outcomes are of different desirability? How do risk and other sociodemographic characteristics affect such interactive decision situations?

During approximately the past two decades, coordination problems, classified in economics as coordination “games” because of their interactive nature, have become an important research topic. Many researchers have tried to answer the preceding and many related micro- and macroeconomics questions through laboratory studies of coordination problems (for relevant surveys and critical assessments of the literature, see Camerer, 2003; Devetag & Ortmann, 2007; Heinemann, 2012; Ochs, 1995).

In this chapter, we discuss four major classes of coordination games in order to convey to our readers the diversity of coordination problems analyzed by economists. We then discuss the design of typical coordination experiments and the experimental protocols that economists use. Then we review “classic” implementations of exemplars of the four classes of coordination games. We conclude with an assessment of the literature: we discuss what we have learned so far from the literature as well as what we consider particularly promising avenues of future research.

## II BACKGROUND AND DEVELOPMENT

We consider four major classes of coordination games.

### A Pure Coordination Games (Rendezvous Games)

Pure coordination games differ from social dilemmas in a key aspect: in social dilemmas, players' preferences over the game outcomes conflict, whereas in pure coordination games players' preferences over outcomes coincide. Players in coordination games, however, are uncertain about the behavior of the other players. This induces what is now in economics called "strategic uncertainty."

The prototypical example of a pure coordination problem, occasionally also called a matching game, is illustrated in the "payoff" table in [Table 16.1](#), where the numbers represent, for example, earnings that experimental subjects might be able to obtain as a function of their choices.

In the matching game in [Table 16.1](#), players have to choose simultaneously, and symmetrically, between two available actions, A and B. The players receive positive, and in fact identical ("payoff-equivalent"), payoffs if they choose the same action, and they both get payoffs of zero if their action choices differ. (In each pair of payoffs, the first number denotes the payoff of the row player, and the second number denotes the payoff of the column player.) The game has two strict Nash equilibria (i.e., equilibria that are robust against slight trembles).<sup>1</sup> Although the actors in the game will be indifferent about which of the equilibria they end up with, they face a problem of coordinating their action choices so as to avoid a disequilibrium outcome that would pay less than either of the strict equilibria. The matching game exemplifies the most fundamental, but by no means trivial, coordination problem. Clearly, the game can be complicated by increasing the dimension of the action space.

Importantly, any preplay agreement that would allow players to coordinate on one of the equilibria is self-enforcing: once players have agreed on playing,

**TABLE 16.1**

		Other player's choice	
		A	B
Your choice	A	1,1	0,0
	B	0,0	1,1

1. Throughout the chapter, we focus on equilibria supported by pure strategies and ignore (the quite possibly numerous) mixed-strategy equilibria. The game shown in [Table 16.1](#) has one mixed-strategy equilibrium that pays (in expectations) less than either of the pure-strategy equilibria.

for example, action A, no one has an incentive to deviate from such an agreement. Apart from communication, which would solve the matching game in a trivial way, anything that allows players to “break the symmetry” between the two equilibria (and make this common knowledge) can function as a coordination device. Symmetry can also be broken when one of the two (or several) possible equilibria is a “focal point.”

Schelling gave more than half a dozen examples of focal points in his widely cited book, *The Strategy of Conflict* (Schelling, 1960). In a famous example, Schelling asked his students to imagine they had to meet someone in New York City without having had a chance to specify a place and time but knowing the other person was also trying to meet them. Where and when would they show up? A large majority of students indicated Grand Central Station at noon as an answer. A simple classroom experiment of the matching game shown in Table 16.1 invariably leads to high degrees of coordination in the upper left corner, where one of the strict equilibria resides, in our experience often with more than 90% of the subjects choosing option A. It is interesting to note, especially in light of the intriguing results of Engelmann and Normann (2010) for a game with Pareto-ranked equilibria, that this happens even in as culturally diverse a society as Australia.

From a game-theoretic standpoint, both “Grand Central at noon” and “option A” can be seen as labels attached to the strategies of a matching game. These labels stand out as unique or more salient than others by virtue of psychological, historical, perceptual, cognitive, or possibly linguistic factors that, of course, beg for explanation.<sup>2</sup> Focal points are the equilibria that result from choosing strategies with salient labels.

## B Pareto-Ranked Coordination Games (Stag-Hunt Games, Order-Statistic Games)

Pareto-ranked equilibria also do not entail, in principle, players’ competing preference orderings over the game outcomes and, like pure coordination games, represent a problem of “strategic uncertainty.” A prototypical example of a Pareto-ranked coordination problem is the stag-hunt game, denoted here as  $g(1,0,x,x)$ , where  $0 < x < 1$ . An instance of such a game is shown in the Table 16.2.

Players again have to choose between two available actions, A and B. The situation can be conceptualized as a weak-link or team effort game in which two players exert either a standard effort independently from one another (hunting hares on their own) or a joint extra effort (hunting a stag together).<sup>3</sup> They both

---

2. A number of recent articles have used “we-thinking”/“team reasoning” to explain these types of results; some have appeared in prominent economics journals (e.g., Bardsley, Mehta, Starmer, & Sugden, 2010). We elaborate on this approach later.

3. The game derives its name, “stag-hunt,” from a parable contained in Rousseau’s writings about the social contract.

**TABLE 16.2**

		Other player's choice	
		<i>A</i>	<i>B</i>
Your choice	<i>A</i>	1,1	0, <i>X</i>
<i>A</i>		X,0	X,X
<i>B</i>			

receive a positive payoff if they choose the same action; furthermore, the payoff from the joint stag-hunt is higher than the one from hunting hares alone. This game, too, has two strict Nash equilibria; however, rather than being payoff-equivalent, these equilibria are Pareto-ranked, with one of the payoffs (resulting from the choice of *A* by both players) being higher for both players than the other (resulting from the choice of *B* by both players).

The stag-hunt game, at first glance, seems to exemplify another fundamental, but more trivial, coordination problem. It also seems that preplay agreements should again be self-enforcing. Or should they?

Assume players have agreed on playing action *A*: does anyone have an incentive to deviate from the agreement? Unfortunately (and this is at the heart of some of the most intriguing results in the literature on coordination), yes: action *B* yields a positive payoff to the player who chooses it *no matter what the other player does*, whereas action *A* is more risky in that it can lead to a payoff of zero to the player who chooses it if he or she is the only one doing so. Hence, the secure action *B* can undermine the risky action *A*, which therefore undermines the possibility of the Pareto-efficient outcome of both players choosing *A*. In fact, even if players agree on choosing *A* prior to playing the game, subsequently they might be tempted to pick *B* if they do not trust the other player (whereas in the matching game, this risk is not present). Intuitively, such a change of mind is also a function of the value that *x* takes. If *x* is greater than 0.5, then the secure action becomes “too attractive”; if *x* is less than 0.5, the secure action turns out to be “not attractive enough.”

Note that the concept of “security” does not distinguish these two cases. If subjects have concerns for security only, then they would always select *B* whether *x* is greater or less than 0.5. This contradicts intuition, which is captured by an alternative solution concept called “risk dominance.” Risk dominance compares the product of the deviation losses of the two equilibria,<sup>4</sup> and it formalizes our intuition that safe actions that are too attractive may induce people to select *B*, whereas safe actions that are not attractive enough might

4. A risk-dominant equilibrium has a greater Nash product of deviation losses relative to the efficient equilibrium (e.g., Harsanyi & Selten, 1988).

induce people to select A. Risk dominance therefore makes the prediction that for high values of  $x$ , the inferior action profile (B,B) will be selected, whereas for low values of  $x$ , the superior, and Pareto-efficient, action profile (A,A) will be selected.

Stag-hunt games are the basic building block of “global games” (Basteck, Daniels, & Heinemann, 2013; Carlson & Van Damme, 1993; Heinemann, 2012; Morris & Shin, 2003). “Global games” are essentially stag-hunt games with payoff-perturbances. This literature has gained some notoriety because it suggests—initially theoretically and recently also experimentally—that risk dominance has more predictive power than payoff dominance.

Stag-hunt games are also the simplest exemplars of so-called order-statistic games. Order-statistic games tend to feature payoff matrices of higher dimension and more than two Pareto-ranked equilibria. They otherwise feature the same tension between a secure action and riskier actions that have the potential for higher (but also lower) payoffs.

Arguably the most prominent example of an order-statistic game is the minimum-effort or weak-link game. In this game, the outcome is determined by the minimum effort exerted (the “weak link” of the chain); any extra effort above such a minimum is costly for the player who exerts it without having any positive effect on the output. Hence, all players gain if the minimum is the highest possible, but each player individually has the incentive not to choose an effort above the minimum. This feature characterizes many team production situations. Respecting deadlines for parts of projects, such as chapters of books, is an appropriate example: all contributors prefer the book to be published as soon as possible, but no one has an incentive to put in an extra effort to respect the deadline if he or she expects that at least one other contributor will be late (Camerer, 2003; Camerer & Knez, 1997).

The payoff function of a generic order-statistic game is  $\pi_i = f(OS - le_i - OSI)$ , where OS is the order statistic chosen (which could be the median or the minimum—the weak link—or something else),  $e_i$  denotes the effort choice,  $le_i - OSI$  denotes the (symmetric) deviation cost, and  $f$  is some scalar function of these terms. Obviously, the terms can be arbitrarily modified by setting the coefficients of the two terms on the right-hand side not equal to 1, by squaring the second term, or by defining the deviation costs asymmetrically. We give more specific examples later.

## C Mixed-Motives Coordination Games (“Battle of the Sexes” Games)

Whereas in pure coordination games players have the same preference ordering over the game outcomes, in mixed-motives games the players differ with respect to the equilibrium they prefer (although they prefer ending up on the equilibrium they dislike to reaching a disequilibrium outcome). The prototypical mixed-motives coordination game is called the “battle of the sexes” and is represented in Table 16.3.

**TABLE 16.3**

		Other player's choice	
		<i>Theater</i>	<i>Football</i>
Your choice			
Theater		2,1	0,0
Football		0,0	1,2

The game takes its name from the example of a couple that has to decide which activity to attend: either going to a football game or to the theater. The husband prefers the theater, whereas the wife prefers the football game, but they both prefer attending either activity together over going to their preferred event alone. The game has two Nash equilibria in pure strategies represented by the two action profile combinations (theater, theater) and (football, football).

Firms often play games that are isomorph to the battle-of-the-sexes game captured in Table 16.3. For example, when two firms engage in collaboration agreements to standardize technologies, both are better off if a single standard develops in the industry, but at the same time each firm prefers its own standard to prevail over that of the rival (Farrell & Saloner, 1986; Stango, 2004). The game also captures situations in which firms have incentives to cooperate to create new business opportunities but in doing so they compete to get the major part of the profits that the new business generates (Brandenburger & Nalebuff, 1997). For example, hardware and software producers have an obvious interest in producing highly complementary technologies to maximize product penetration in the market; at the same time, each wants to extract the largest share of the profits generated by the collaboration.

When the game is played only once, chances are high that a disequilibrium outcome occurs; preplay communication—either one- or two-sided—might facilitate the chances of reaching equilibrium. If the positions of the two players are asymmetric, the one with higher bargaining power has an advantage. If the game is played repeatedly, a reasonable (and fair) pattern of coordination is to alternate between equilibria.

## D Critical-Mass Games (Panics, Revolutions)

Equilibria in a coordination game can be interpreted as conventions that arise within a population so that no one has an incentive to deviate from a conventional behavior if everybody else adheres to it. Sometimes the emergence of specific conventions depends on a critical number of adopters. If people who start behaving in a certain way are numerous enough, that behavior will eventually become a convention; otherwise, it will disappear or remain customary only within a minority. Schelling (1978) mentioned several examples of “critical-mass” effects in the social sciences, from fashion to panic crises and revolutions.

Critical-mass effects are also important in economics. For example, they are important in understanding the adoption of technologies or products subject to “network externalities” (where the incentive to adopt a technology increases as the number of other adopters of the technology increases).<sup>5</sup> The large-scale diffusion of these technologies often depends on whether the “installed base” of early adopters is large enough to generate a bandwagon effect (e.g., [Farrell & Saloner, 1986](#)).

The following is an example of a critical-mass game with binary choices:  $N$  players simultaneously must decide whether to invest a certain amount in a network technology that is installed if the total revenue is at least  $ZK$ , where  $1 < K \leq N$ . Contribution yields a return greater than  $Z$  if and only if at least  $K$  players contribute ([Heinemann, Nagel, & Ockenfels, 2004](#); for a dynamic version, see [Duffy & Ochs, 2012](#)). In some cases, different conventions are possible, some of which yield greater returns than others but require a higher critical number of adopters.

An example of a critical mass game, with  $N=7$  possible choices and with  $N=7$  players, is shown in [Table 16.4](#).

Each of the  $N$  players can choose between seven actions (e.g., behaviors, products, and technologies); higher-numbered actions yield greater returns but are also riskier because they require higher critical numbers of adopters in order to yield positive payoffs. There are seven equilibria corresponding to all players choosing the same action. The equilibria are Pareto-ranked, and higher-payoff equilibria are riskier than lower-payoff equilibria. As in stag-hunt and order-statistic games, there is a trade-off between efficiency and security.

**TABLE 16.4**

Your choice of X	No. of people who have chosen X						
	7	6	5	4	3	2	1
7	7	0	0	0	0	0	0
6	6	6	0	0	0	0	0
5	5	5	5	0	0	0	0
4	4	4	4	4	0	0	0
3	3	3	3	3	3	0	0
2	2	2	2	2	2	2	0
1	1	1	1	1	1	1	1

5. All communication technologies (e.g., fax, e-mail, and mobile phones) are subject to network externalities: the utility for any adopter increases with the number of other adopters.

Unlike order-statistic games, however, payoffs depend on absolute frequencies of people choosing certain actions and not on “minimum” or “median” players.

Other real-world examples are processes of organizational change: low-complementarity norms and routines (e.g., regarding division of labor among divisions, offices, or co-workers) usually provide a high level of redundancy against “breakdowns” but are relatively inefficient; more efficient norms may require the effective contribution of more (or all) organization members in order to be effective. A related example is the gradual emergence of a new norm or behavioral “standard,” which may have to be adopted by a large proportion of organizational members to lead to improvements.

Yet other examples of which the game in [Table 16.4](#) captures the essence are choices between, for example, activities that are more or less enjoyable but some of which cannot be performed unless enough people adhere, where “enough” has a different value for each activity. One can rent a movie and watch it alone at home, but one needs a companion to play chess; at the same time, both would prefer going to a party, as long as everybody else will go. Schelling mentions the example of the last day of a class when some students hesitantly begin clapping as the teacher prepares to leave the room: “If enough clap, the whole class may break into applause; if a few clap indecisively, it dwindles to an embarrassed silence” ([Schelling, 1978, p. 90](#)). An application of these ideas to financial crises of various kinds may be found in [Heinemann \(2012\)](#).

### III DESIGN AND IMPLEMENTATION OF COORDINATION EXPERIMENTS IN ECONOMICS

The design and implementation of economics experiments follows, by far, a well-established canon of methodological precepts ([Hertwig & Ortmann, 2001](#)). One of the key precepts of experimental economics is that deception should not be used. Essentially, experimental economists are concerned that deception could lead to negative reputational spillover effects that might contaminate their subject pool in future experiments. Indeed, suspected deception has been shown to provoke a range of motivational, emotional, and cognitive reactions on the part of subjects (for a review of the evidence from psychology, in which deception is still often used, see [Hertwig & Ortmann, 2008](#); [Ortmann & Hertwig, 2002](#)). There are also reasons to believe that deception and the reaction it causes lead to important changes in the readily available subject pools. Obviously, suspicion of deception could immensely affect subjects’ choices in coordination games. Therefore, the de facto proscription of deception in experimental economics—top economics journals refuse to publish results of experiments in which deception was used—is a crucial step toward experiments with some external validity.

Another relevant tenet of economic experiments is that participants in experiments role-act. They are made to take on, for example, the role of a buyer and seller in a transaction of a good or service or the role of a decision-maker

in a coordination problem. Until very recently, it was argued that the specific situation should be presented to subjects in abstract (and therefore allegedly neutral) terms: rather than being called a buyer or a seller, the two participants in an experimental transaction would be called participant A and participant B, and they would trade an unspecified good. This procedure was rationalized by the assumption that it would guarantee that subjects would not bring prior expectations (e.g., from previous buyer–seller interactions) into the lab that might contaminate their reactions to experimental stimuli.

In light of important contradictory evidence, this assumption has been questioned, and it has been argued that a better experimental strategy would be to systematically study the impact of context (e.g., [Ortmann & Gigerenzer, 1997](#); see also the important discussion of field experiments in [Harrison & List, 2004](#)). Ortmann and Gigerenzer provide an early review of evidence that suggests that embedding abstract reasoning tasks such as the Wason selection task in thematic garb with which subjects are familiar (i.e., framing this “four-card problem” as a social contract story rather than one involving letters and numbers) can lead to dramatic improvements in performance—here defined as the identification of the “logically” correct answer. These authors have also demonstrated similar context effects for related economic games such as gift-exchange games. To the extent that some classes of coordination games are considered to be appropriate models of teamwork or other forms of coordination within firms, this issue is of obvious importance. Recent attempts by coordination problem researchers to add context to their laboratory settings is therefore another welcome step toward experiments with some external validity. For example, [Bortolotti, Devetag, and Ortmann \(2013\)](#) designed a coordination game experiment in which subjects were engaged in a real task (counting coins under time constraints) in which, therefore, real as opposed to hypothetical effort was involved. Bortolotti et al.’s findings indicate a substantially higher degree of coordination success when subjects acted in teams and were rewarded according to a “minimum effort” payoff function than has been found in comparable “chosen effort” games. It seems that this novel result is due to subjects, being aware that it takes some time to learn the task, giving other subjects the benefit of the doubt when those other subjects do not immediately manage to expend the maximum effort. This seems to be a more “realistic” reflection of what happens in teams. Further “field” experiments of this sort in the domain of coordination games seem warranted in order to increase the external validity of laboratory findings.

The issue of financial incentives is probably the most prominent characteristic of economics experiments. The term “financial incentives” denotes, typically, monetary rewards that are a function of the action(s) a player chooses as well as the actions other players choose. For example, although subjects might be paid a show-up fee (or participation fee) upfront, a significant part of their earnings comes from “performance-based” payments where performance is typically relative to a game or decision-theoretic benchmark.

This strictly enforced experimental requirement, by making participants' decisions to some extent costly for them (e.g., through forgone earnings), was hypothesized to affect participants' decisions. The available evidence supports this conjecture. After reviewing the available evidence empirically, [Hertwig and Ortmann \(2001\)](#) stated:

*To conclude, concerning the controversial issue of the effects of financial incentives, there seems to be agreement on at least the following points: first, financial incentives matter more in some areas than in others (e.g., see Camerer & Hogarth's distinction between judgment and decision vs. games and markets). Second, they matter more often than not in those areas that we explore here (in particular, research on judgment and decision-making), which are relevant for both psychologists and economists. Third, the obtained effects seemed to be two-fold, namely, convergence of the data toward the performance criterion and reduction of the data's variance. (p. 395)<sup>6</sup>*

A number of studies have confirmed these results (e.g., the study of [Gneezy & Rustichini, 2000](#); see also the reanalysis of these data in [Rydval & Ortmann, 2004](#)). Importantly, in light of the basic structure of many coordination problems—risky but potentially payoff-improving action choices versus safe action choices that, however, pay relatively little—a controversy reported in the *American Economic Review* is noteworthy. [Holt and Laury \(2002, 2005\)](#) showed that increasing the financial incentives for decisions over lotteries (i.e., risky choices with potentially high payoffs), and doing so quite dramatically, increased risk aversion.<sup>7</sup>

By far, the design and implementation of coordination games is straightforward from a methodological standpoint. Subjects are recruited (either through flyers or e-mails) and come to a laboratory where they are seated and presented with stimuli materials. The stimuli materials, in coordination games, almost always contain earnings tables of the kind presented in the preceding section (see [Tables 16.1–16.4](#)). Subjects are then asked, typically repeatedly, to decide which of the available action choices they are willing to choose. How often they are asked turns out to be of some importance, roughly for the reason reported

---

6. [Hertwig and Ortmann \(2001\)](#) also argue, thereby questioning the current practice among experimental economists, that “researchers seeking maximal performance ought to make a decision about appropriate incentives. This decision should be informed by the evidence available.... In cases where there is no or only mixed evidence, we propose that researchers employ a simple ‘do-it-both-ways’ rule. That is, we propose that the different realizations of the key variables discussed here, such as the use or non-use of financial incentives ... be accorded the status of independent variables in the experiments” (p. 400).

7. [Harrison, Johnson McInnes, and Rutström \(2005\)](#) made a persuasive case that order effects confounded the original results and that the effect of increased financial incentives, although an important qualitative phenomenon, was only approximately half of what had been reported in [Holt and Laury \(2002\)](#).

by Holt and Laury (2002): the less that is at stake at each decision (typically that means more rounds overall are being conducted), the less expensive is experimentation for the players (i.e., the choice of the risky but potentially very rewarding action choice), as convincingly shown by Berninghaus and Ehrhart (1998) for games with Pareto-ranked equilibria. Other implementation details that have similar effects are an increase in the number of actions available to subjects, or a reduction in the costs of deviating from equilibrium play (see more details in [Section VI](#)).

Another important design choice in coordination game experiments concerns the interaction protocol: in some studies, each player always interacts with the same opponent(s) throughout the game duration (the so-called “fixed matching” protocol), whereas in other studies subjects play a game repeatedly but change the person they are matched with every round, either by a random scheme or by a deterministic rotation scheme. In games with groups of players (i.e., more than two), the composition of the groups can remain fixed throughout the game, or, at some point, different groups can be “scrambled” randomly to form new groups.

Random rematching is typically introduced to induce subjects to use behavioral strategies that they would use if the game was to be played only once. This design strategy allows one to observe behavior in a series of independent “one-shot” games without employing huge numbers of subjects (of course, the random scheme does not impede learning from experience, but it does impede, for example, coordination based on precedent). The choice of the interaction structure in coordination games is obviously relevant in that some coordination devices can be used by players in the first interaction protocols but not in the second ones, with relevant differences for the possibility of observing coordination failure (this is discussed further in [Section V](#)).

## IV TECHNOLOGICAL DEVELOPMENTS

More than 50 years have passed since the first economics experiments were conducted. During that time, some implementation features have changed as a consequence of developments in technology. Whereas the earlier experiments were conducted with paper and pencil, the norm nowadays is to have experimental subjects interact through networked computers. The advantages of computerized experiments over the traditional paper-and-pencil methods are evident and have allowed the discipline to make enormous progress in a relatively short time span. Games that are repeated for many rounds (e.g., 100 or more) in order to test alternative learning theories would, for example, not be feasible without the aid of computers. Computerized experimental sessions, furthermore, allow the experimenters to keep track of an enormous amount of data well beyond the decision itself (e.g., time responses and patterns of information search).

A disadvantage of computerized experiments is the possibility of a decrease in transparency (and hence a slightly higher probability that participants may think they are being deceived). If, for example, a random device must be used in an experiment, a computerized random number generator is less transparent than the act of throwing a dice in front of participants. Such implementation details have been shown to matter in some circumstances (see [Hertwig & Ortmann, 2001](#)).

The path-breaking experiments on coordination games by [Van Huyck, Battalio, and Beil \(1990, 1991\)](#) and by [Cooper, De Jong, Forsythe, and Ross \(1990, 1992\)](#) were run with paper and pencil, but most subsequent studies were computer-aided experiments. One of these ([Van Huyck, Battalio, & Rankin, 2001](#)) used a graphical interface whereby players could see payoffs corresponding to alternative choices by moving a cursor along the rows and columns of a matrix instead of having an earnings table in paper format. Because the actions available to a subject in this experiment (a variant of the “weak link” game discussed previously) were 100 instead of the 7 used in the early experiments, this implementation option was almost unavoidable, although some doubts remain regarding whether discovering potential payoffs in this manner—as opposed to having them displayed on a table—makes a difference in terms of behavior. No comparative study has been done so far; hence, we can only conjecture that the presentation format may have had an impact on, for example, the different salience of available actions in the first round or on the ease of processing the payoff consequences of out-of-equilibrium choices.

Surprisingly, another well-established computerized methodology to record patterns of information search, MouseLab, has never been used in coordination game experiments.<sup>8</sup> With the MouseLab technology, participants must click on a cell of the earnings table to visualize its content, thereby making it possible to observe which pieces of information are taken into consideration by subjects in deciding which action to take, and in which sequence such information is acquired. In experiments on repeated coordination games, for example, using MouseLab techniques would provide insights on which payoff information drives first-round choices. Another similar technique that could be used with much insight to gain is eye-tracking technology, which allows one to observe what information subjects pay attention to when deciding and in which particular manner or sequence different pieces of information are put in relation to one another (e.g., depending on whether a subject acting as row player looks at his or her payoffs “by row” or “by column,” one can infer with a certain degree of probability whether that subject is performing an average payoff calculation or detecting a dominance relation).

---

8. For examples of experiments on games conducted with MouseLab, see [Johnson, Camerer, Sen, and Rymon \(2002\)](#) and [Costa-Gomes, Crawford, and Broseta \(2001\)](#).

## V EXAMPLES

### A Experimenting with Pure Coordination Games: Mehta, Starmer, and Sugden (1994)

Matching games are especially useful in studying the influence of culture, language, and shared codes on coordination. Standard game theory does not discriminate between payoff-equivalent equilibria, and the usual refinements such as payoff dominance and risk dominance do not apply to matching games. Therefore, matching games are an ideal point of departure for the empirical investigation of how people solve coordination problems. Following Schelling's original intuition on focal points, Mehta et al. (1994) decided to test experimentally the influence of salience on coordination and replicated some of Schelling's early informal experiments in a more rigorous manner.<sup>9</sup>

Mehta et al. (1994) started by providing a definition of different forms of salience. They distinguish between *primary salience*, consisting of an individual's personal preference for a particular option among other available options; *secondary salience*, which is derived from knowing that a particular option has primary salience for the person with whom one has to coordinate; and *Schelling salience*, which is derived only from the probability that a particular option may be selected as the solution to a coordination problem. An option need not be characterized by all three types of salience. For instance, if 9 is my favorite number, such a number will have primary salience for me. Number 5 has secondary salience for me if I know that it is my best friend's favorite number. Furthermore, if my best friend and I play a matching game in which we have to pick the same number to be rewarded, we may pick number 1, which is characterized by high Schelling salience. In this example, three different options have different types of salience. If, instead, we both had 9 as our favorite number and if this information was common knowledge between us, we could pick number 9 in a matching game. In this last case, the three types of salience would select the same option.

The experimental study reported in Mehta et al. (1994) was aimed at verifying whether coordination success in matching games is related to primary salience (which would imply that some options simply tend to be preferred to others by most people) or to Schelling salience. For this purpose, they divided their pool of subjects into two groups. In one group (the "coordinating" group), subjects were simply asked to select from a set of options the one they preferred. Subjects in the other group (the "picking" group) were asked to select an option in the same set knowing that they would be paired with another participant randomly and both would earn a prize if the option selected was

---

9. That is, by providing financial incentives to subjects and by preventing any form of communication, in the typical tradition of experimental economics experiments (see the preceding section for a discussion).

the same. Thus, subjects in the second group were playing a matching game. The options were taken from several sets, including dates, flowers, cities, and male names.

The results clearly indicate the existence of Schelling salience, which is often uncorrelated with the other types of salience.<sup>10</sup> For example, when asked to specify a date, subjects indicated 75 different dates; however, in the “coordinating” group 50% of subjects indicated December 25. Analogously, when asked to specify a male name, only 9% of subjects indicated John in the “picking” condition, but 50% did so in the “coordinating” condition. These experiments show that people are generally good at solving coordination problems. Coordination success is not related to the fact that some options are more preferred than others<sup>11</sup> but, rather, to the fact that people within a population tend to agree on which options are unique and distinctive regardless of individual preferences (which are often highly heterogeneous). The results, of course, beg an answer to the questions of why people are so good at solving pure coordination games in certain situations and how Schelling salience comes into existence. With respect to these questions, recent contributions bring into play a phenomenon called “we-thinking” or “team reasoning.” Team reasoning (e.g., [Bardsley et al., 2010](#)) has been defined as a decision criterion based on collective rather than individual rationality. A player asks him- or herself not “What do I want and what should I do to achieve it?” but, rather, “What do we want, and what should I do to help achieve it?” ([Colman, Pulford, & Rose, 2008, p. 389](#)). According to Colman et al., team reasoning supports the selection of focal points in coordination games. We are skeptical about the explanatory power of this approach, which seems not to be able to handle the well-documented sensitivity of coordination games to various design and implementation details such as number of players, stakes, or other manipulations that could affect strategic uncertainty (e.g., [Devetag & Ortmann, 2007](#)). Also, the basic assumption of people engaging in we-thinking seems not well-established (and is surely moderated by group identity), and it seems conditional on design and implementation details that might affect the degree of strategic uncertainty as well.

## B Experimenting with Mixed-Motives Games: [Cooper, De Jong, Forsythe, and Ross \(1989\)](#)

In the “battle-of-the-sexes” game described previously, players’ conflict of preferences over the two equilibria is dominated by their common incentive

---

10. [Harrison \(2005\)](#) argues, and makes a persuasive case empirically, that at least some treatments were confounded by natural language: “There is a fundamental confound in experiments such as these: the fact that natural language has been used to present the task to subjects, and that the task itself uses natural language. That language itself has salient labels, which is just to say that some words are prominent or conspicuous” (p. 21).

11. Primary salience and Schelling salience were usually unrelated in the experiment.

**TABLE 16.5**

		Other player's choice	
Your choice		1	2
1	0,0	200,600	
2	600,200	0,0	

to coordinate. Path-breaking experiments on mixed-motive games of coordination were run by [Cooper et al. \(1989\)](#), who used the payoff matrix in [Table 16.5](#).

The Nash equilibria in pure strategies correspond to the two action profiles (1,2) and (2,1). (The game also has an equilibrium in mixed strategies in which players play strategy 1 with probability 0.25 and strategy 2 with probability 0.75. The expected payoff of the mixed-strategy Nash equilibrium is 150.) Without a device that would allow breaking the symmetry between players and without the possibility to communicate, the chances that a disequilibrium outcome occurs are high. Indeed, [Cooper et al. \(1989\)](#) report data of an experiment in which cohorts of 11 players were playing a series of 22 one-shot games like the one shown in [Table 16.5](#), with different anonymous opponents and with players alternating between the role of row and column player. More than half of the rounds result in a disequilibrium outcome, suggesting that subjects were either not able or not willing to capture the significant gains from coordination.

Given this baseline result, [Cooper et al. \(1989\)](#) then tested the effect of communication on the chances of coordination; in doing so, they distinguished between one-way and two-way communication. In the one-way communication treatment, the row player can send the opponent a nonbinding, costless message on his or her choice intentions prior to playing the game. The column player cannot respond. In the two-way communication treatment, both players can send each other costless and nonbinding messages declaring what they intend to play. Messages are nonbinding in the sense that a player is not constrained to actually choose according to what he or she has announced (“cheap talk”). After messages have been sent, pairs of subjects play out the game.

For one-way communication, a plausible outcome is for the player who can send the message to declare his intention to play the strategy supporting his preferred equilibrium. If the opponent considers the announcement credible enough, she will best respond to the announced strategy. This way, the player who can communicate has the advantage of favoring coordination on his preferred equilibrium, and both players are better off, on average, than without the announcement. [Cooper et al.'s \(1989\)](#) data show that, in fact, one-way communication improves coordination dramatically and favors, in the majority of cases, the player that had the possibility to send the message. Equilibrium play in the

one-way communication treatment is observed 95% of the time, and in the large majority of cases row players announce their intention to play strategy 2 and actually play it. Column players, on their part, almost always play strategy 1. Two-way communication, in contrast, makes coordination *less* likely compared to one-way communication and almost as likely as coordination without communication. In fact, in the majority of cases, both players announce their intention to play the strategy implementing their preferred equilibrium, thus reintroducing the perfect symmetry that one-way communication had broken. As a result, although 80% of equilibrium announcements result in equilibrium play, 45% of rounds result in an outcome of disequilibrium.

### C Experimenting with Order-Statistic and Stag-Hunt Games: Van Huyck et al. (1990, 1991), Cooper et al. (1990, 1992), and Rankin, Van Huyck, and Battalio (2000)

Van Huyck et al. (1990, 1991) used the earnings tables shown in Tables 16.6 and 16.7 for their two path-breaking studies. These studies are particular implementations of the order statistic games mentioned previously.<sup>12</sup>

In both experiments, subjects had to choose numbers between 1 and 7. In the median game, everybody's payoff increases in the median of all numbers and decreases with the distance between the number chosen and the median. In the minimum or "weak link" game described previously, everybody's payoff increases in the minimum of all numbers chosen and again decreases with the distance in the number chosen and the minimum. Note that although the Pareto-ranked equilibria on the main diagonal are the same, the off-diagonal elements differ. That is because they were generated by slightly different payoff functions.

The payoff-dominant or efficient equilibrium is in the upper left corner for both the minimum game and the median game, whereas the secure action induces an equilibrium (the secure equilibrium from here on) in the lower right corner for the minimum game and two rows up from the bottom in the median game. Both games feature seven (identical) Pareto-ranked strict equilibria on the main diagonal. There is a tension between the secure action—the lowest action in the minimum game and the third lowest in the median game—and the action required for efficient equilibrium. If payoff dominance, or efficiency, was the focal equilibrium, as intuited by almost all economists initially and also suggested by Harsanyi and Selten (1988), then subjects should have selected the strategy with the highest number.

Was efficiency psychologically salient in Van Huyck et al. (1990, 1991) or were competing concepts such as security, or risk dominance, more salient? The key result of their 1990 work is the stable and speedy unraveling of action

---

12. The results of these studies are, in our view rightly so, among the most celebrated in the literature on coordination failure (e.g., Camerer, 2003; Ochs, 1995; <http://scholar.google.com>).

**TABLE 16.6** Earnings Table for the “Median Game”

	Median value of X chosen							
Your choice of X	7	6	5	4	3	2	1	
7	1.30	1.15	0.90	0.55	0.10	-0.45	-1.10	
6	1.25	1.20	1.05	0.8	0.45	0.00	-0.55	
5	1.10	1.15	1.10	0.95	0.70	0.35	-0.10	
4	0.85	1.00	1.05	1.00	0.85	0.60	0.25	
3	0.50	0.75	0.90	0.95	0.90	0.75	0.50	
2	0.05	0.40	0.65	0.80	0.85	0.80	0.65	
1	-0.5	-0.05	0.3	0.55	0.70	0.75	0.70	

**Source:** From Table  $\gamma$  in Van Huyck et al. (1991).

**TABLE 16.7** Earnings Table for the “Minimum Game”

Your choice of X	Smallest value of X chosen							
	7	6	5	4	3	2	1	
7	1.30	1.10	0.90	0.70	0.50	0.30	0.10	
6	—	1.20	1.00	0.80	0.60	0.40	0.20	
5	—	—	1.10	0.90	0.70	0.50	0.30	
4	—	—	—	1.00	0.80	0.60	0.40	
3	—	—	—	—	0.90	0.70	0.50	
2	—	—	—	—	—	0.80	0.60	
1	—	—	—	—	—	—	0.70	

Source: From Table A in Van Huyck et al. (1990).

choices to the worst of the strict equilibria. In the treatments based on [Table 16.7](#), 14–16 participants played the stage game repeatedly (10 times in treatment A and 5 times in treatment A') and for money, receiving information about their payoffs only after each stage. The outcome was essentially the same even after payoff-efficient precedents emerged in a treatment (B) that was inserted between treatments A and A' for four out of six sessions. Several other researchers (with baseline treatments for various modifications reported in those papers) replicated this unraveling result with the same payoff matrix and with subject numbers varying from 6 to 14. Other experimenters (also with baseline treatments for various modifications reported in those papers) chose structurally similar payoff matrices (e.g., linear deviation costs and no negative payoffs) with slightly more or less action choices and also replicated this result. The detailed references may be found in [Devetag and Ortmann \(2007\)](#).

Note that the action choices are hypothetical, although they might have tangible consequences. In this sense effort, for example, is “chosen” and not real. We return to this issue later.

[Van Huyck et al. \(1991\)](#) demonstrate the influential role of the initial action choices. For the baseline treatment, neither the unique payoff-dominant equilibrium nor the unique secure equilibrium emerged when nine participants played the stage game repeatedly (again, 10 times), receiving information about their payoffs only after each stage. The initial median constituted a strong precedent from which subjects had trouble extracting themselves. This result, too, has been replicated by other authors. The detailed references may be found in [Devetag and Ortmann \(2007\)](#).

Not surprisingly, the remarkable coordination failure results of [Van Huyck et al. \(1990, 1991\)](#) (produced under conditions that, on the surface, seemed rather conducive to coordination successes, e.g., small groups, perfect information, and easy-to-understand task) drew considerable attention and a steady flow of attempts to test their robustness.<sup>13</sup> We discuss some of these attempts in [Section VI](#).

The basic stag-hunt game explored experimentally by [Cooper et al. \(1992\)](#) was  $1,000g(1,0,0.8,0.8)$  and is represented in [Table 16.8](#).

As in the order-statistic games discussed previously, the payoff-dominant equilibrium is in the upper left corner while the secure equilibrium is in the lower right corner. There is thus again a tension between A, the risky action (that might induce the efficient equilibrium), and the secure action (that induces

---

13. [Van Huyck et al. \(1990, 1991\)](#) conducted a number of important robustness tests. Among their key insights are the importance of the number of participants, the matching protocol, the feedback conditions, and the deviation cost. In [Van Huyck et al. \(1990\)](#), for example, the authors demonstrated (in the already mentioned treatment B) that setting the coefficient on the deviation cost equal to zero led to quick convergence to efficiency. They also demonstrated that two participants when matched repeatedly and with the same person (but not with randomly drawn others) were able to coordinate on the efficient outcome.

**TABLE 16.8**

		Other player's choice	
Your choice	A	B	
A	1,000,1,000	0,800	
B	800,0	800,800	

the inefficient equilibrium, or an outcome even worse for the other player). The same kind of stag-hunt game was also the basic building block of Cooper et al. (1990). A key question that these researchers asked in both of their articles was whether the Pareto-dominant equilibrium would always be selected. The answer to this question was negative.

Following up on related work published in 1989, Cooper et al. (1992) also explored whether the coordination failure results in their earlier work (1990) were robust to the use of both one-way and two-way communication, for this particular parameterization of the stag-hunt game. Coordination failure turned out to be endemic in the no-communication baseline conditions (and still significant with one-way communication); coordination failure was eliminated by two-way communication between players.

It is important to mention that the coordination failure results of Cooper et al. (1990, 1992) came about under a matching protocol that differed sharply from the one used by Van Huyck et al. (1990, 1991) and other multiplayer studies afterwards. Specifically, whereas Van Huyck et al. routinely used multiplayer, finitely repeated coordination games with fixed matching, Cooper et al. (1989, 1990, 1992) used two-player sequences of one-shot games resulting from a random matching or rotation matching (Kamecke, 1997) protocol. As was shown later by, for example, Clark and Sefton (2001), the choice of the latter interaction patterns makes an efficiency-reducing difference.<sup>14</sup> It has also been shown, for example, that in a multiplayers repeated game whereby group composition remains fixed over time, some players are willing to choose efficient actions in initial rounds, and by doing so they forego higher initial payoffs with the purpose of “teaching” the other players to do the same and converge to the efficient equilibrium eventually. This behavior obviously is not applicable under a random rematching scheme.

---

14. The authors had their subjects participate in a stag-hunt game either as a sequence of one-shot games implying a random matching protocol or as a repeated game with a fixed matching protocol. Their data show that, indeed, in the first round of play, the frequencies of choice of the risky action were 0.3 in the random matching and 0.6 in the fixed matching protocol, a highly significant difference. Moreover, the fixed matching protocol reduced the instances of disequilibrium outcomes and increased the overall proportion of risky choices across rounds.

Arguably the most intriguing article in this area is [Rankin et al. \(2000\)](#). The authors use a scaled-up version of  $g(1,0,x,x)$ , where  $x$  is, for each round, drawn randomly from the unit-interval and then, ever so slightly, perturbed. Rankin et al. had their subjects play a sequence of 75 such games, in addition scrambling the action labels so that the payoff-dominant equilibrium and the secure equilibrium would not show up in the same corner throughout the 75 rounds. The intriguing result of this experiment—which was explicitly motivated by an attempt to increase external validity—was the high percentage of efficient play both when  $x < 0.5$  (making the secure strategy less attractive and making payoff-dominant and risk-dominant equilibrium coincide) and when  $x > 0.5$  (making the secure strategy more attractive and positioning the payoff-dominant and the risk-dominant equilibrium at opposite ends of the main diagonal).<sup>15</sup>

[Rankin et al. \(2000\)](#) note that their setup inhibits learning from experience and focuses subjects on the exploration of deductive principles. In addition, in approximately half of the rounds, subjects faced a situation in which payoff dominance and risk dominance selected the same equilibrium. Obviously, the results these researcher have reported are dramatically at odds with claims that coordination failure is common.

## D Experimenting with Critical-Mass Games: [Devetag \(2003\)](#)

[Devetag \(2003\)](#) studied coordination in a repeated critical-mass game with seven strategies. The baseline version of the critical-mass game was shown in [Table 16.4](#) and, for integers in a range  $[1, \dots, I]$ , is generated by the following payoff function:

$$\pi_i = \begin{cases} i & \text{if } k_i \geq i \\ 0 & \text{otherwise} \end{cases}$$

where  $i$  is the integer chosen by a player, and  $k_i$  is the total number of players in the groups who have chosen that integer. A version of the critical-mass game featuring increasing returns can be generated by the following payoff function:

$$\pi_i = \begin{cases} i + k_i - 1 & \text{if } k_i \geq i \\ 0 & \text{otherwise} \end{cases}$$

In the second payoff function, given that the threshold for a number is matched, the individual payoff is higher the higher the number of players that choose it. Theoretically and anecdotally, critical-mass effects tend to be associated with increasing returns, implying that once a threshold is reached, the more

---

15. Specifically, for the first 10 periods, 65% (85%) of choices corresponded to the efficient action when  $x > 0.5$  ( $x < 0.5$ ). For the last 10 periods, approximately 90% (~100%) of the choices corresponded to the efficient action when  $x > 0.5$  ( $x < 0.5$ ). Thus, payoff dominance clearly carried the day.

individuals engage in some activity, the more others will be inclined to do the same. If the number of players  $N$  is set equal to the number of available choices  $I$ , the attainment of the highest payoff implies that all players in the group must pick the highest integer.

[Devetag \(2003\)](#) investigated coordination in the critical-mass game both with and without increasing returns. Groups of seven subjects played one of the two games for 14 periods with a fixed matching protocol. The main treatment variable was the information condition. In the full information condition, players knew the distribution of all choices after each round. In this information condition, coordination on the payoff-dominant equilibrium was the most frequent outcome, and such outcome was even more frequent when groups played the version of the game with increasing returns. The path of play was often a coordinated, one-step-at-a-time movement toward the payoff-dominant equilibrium, sometimes allowing subjects to escape even from inefficient equilibria. Such “creeping up” did not occur when subjects only knew the median of all choices. Players did use this information to coordinate on the historical median, but this was an inefficient median in all groups. Finally, no convergence, even on suboptimal equilibria, was observed when players only had information about their own payoff. Hence, average efficiency increased monotonically with the increase in information, although only full feedback allowed convergence to the best equilibrium.

The individual behavior analysis reveals that the full feedback condition was actively used by some players (the “leaders”) to signal to others the choice of the efficient equilibrium, inducing in most cases the remaining players to eventually pick the efficient action.

Semeshenko et al. (2010) studied a similar critical-mass coordination game with heterogeneous individuals under different information treatments. They explored the effect of a gradual decrease in information on the emergence of Pareto-efficient outcomes. They observed that successful coordination is possible even with private information, albeit not on a Pareto-optimal equilibrium.

## VI ASSESSMENT

The early results of [Van Huyck et al. \(1990, 1991\)](#) and [Cooper et al. \(1990, 1992\)](#) seemed to indicate—somewhat in contrast to the results on pure coordination games (e.g., [Mehta et al., 1994](#)), critical-mass games ([Devetag, 2003](#)), and also market entry games (e.g., and famously, [Kahneman, 1988<sup>16</sup>](#))—that

---

16. [Kahneman \(1988\)](#) had  $N$  participants choose simultaneously, and without communicating, whether to enter a market or not. The market had a “carrying capacity” of firms that the market could sustain. Entry in excess of the carrying ended in losses for all participants, whereas lack of entry led to forgone welfare gains. The results of his experiment (later prominently used in [Camerer & Lovallo, 1999](#)) stunned Kahneman, who famously said, “To a psychologist, it looks like magic.” A good discussion of market entry games may be found in [Camerer \(2003\)](#).

coordination failure, at least for order-statistic games and stag-hunt games, was a common phenomenon. Not surprisingly, then, a number of authors followed up on these studies. For stag-hunt games, we already discussed the remarkable results of [Rankin et al. \(2000\)](#) that suggested strongly that payoff dominance had been counted out as a selection principle too early. Other related work is discussed in [Camerer \(2003\)](#) and [Devetag and Ortmann \(2007\)](#).

The order-statistic games by [Van Huyck et al. \(1990, 1991\)](#), especially the minimum game, have recently seen renewed interest from top economics journals (e.g., [Brandts & Cooper, 2006a](#); [Blume & Ortmann, 2007](#); [Chaudhuri, Schotter, & Sopher, 2009](#); [Weber, 2006](#)). What has made the minimum game so attractive is the claim, not undisputed, that it is a good model of teamwork and therefore a basic building block of organizations—a theme already explored in the early work by [Camerer and Knez \(1996, 1997\)](#). In light of that claim, the initial evidence of coordination failure being pervasive was almost bound to lead to attempts to test experimentally the robustness of this evidence.

In [Devetag and Ortmann \(2007\)](#), we discussed the myriad of ways through which researchers have explored whether coordination failure is indeed a robust phenomenon. We documented there that the original results have indeed been shown to be robust in that they have been easy to replicate. We have, however, also documented what can be done to engineer coordination successes. It is now, for example, well-established that lowering the attractiveness of the secure action relative to the risky action that is required to implement the efficient equilibrium (e.g., [Brandts & Cooper, 2006a](#)) is an efficiency-enhancing design choice. Likewise, lowering the costs (expressed as foregone payoffs) that subjects suffer from trying riskier but potentially more rewarding strategies (“experimentation”) has been shown to work quite well. This has been achieved by lowering the payoffs associated with out-of-equilibrium choices *ceteris paribus* (e.g., [Battalio, Samuelson, & Van Huyck, 2001](#); [Goeree & Holt, 2005](#); [Van Huyck et al., 1990](#)), by increasing the number of rounds while keeping the overall earnings roughly the same (e.g., [Berninghaus & Ehrhart, 1998](#)), or by refining the action grid while keeping the range of available payoffs the same ([Van Huyck et al., 2001](#)).

In addition, fixed matching protocols promote efficiency (e.g., [Clark & Sefton, 2001](#); [Schmidt, Shupp, Walker, & Ostrom, 2003](#); [Van Huyck et al., 1990](#)). Even random matching schemes can favor efficiency if the experimental design and implementation induces subjects to focus on the deductive principles underlying the game rather than on one’s own payoff history associated with different actions (e.g., [Rankin et al., 2000](#); see also [Schmidt et al., 2003](#)). Providing full informational feedback seems efficiency enhancing in “small” groups (e.g., [Berninghaus & Ehrhart, 2001](#); [Brandts & Cooper, 2006b](#); [Weber, 2006](#); but see [Devetag, 2005](#)), and the possibility of observing other players’ expressions of intent and subsequent action choices was also shown to increase efficiency ([Duffy & Feltovich, 2002, 2006](#)). Finally, both costly (e.g., [Cachon & Camerer, 1996](#); [Van Huyck, Battalio, & Beil, 1993](#)) and costless preplay

communication are efficiency-enhancing devices (e.g., Bangun, Chaudhuri, Prak, & Zhou, 2006; Blume & Ortmann, 2007; Cooper et al., 1992; Duffy & Feltovich, 2002, 2006; Van Huyck, Gillette, & Battalio, 1992), as is higher quality of information, when this is made common knowledge (Chaudhuri et al., 2009; see also Bangun et al., 2006).

In Devetag and Ortmann (2007), we argued that many of the strategies to engineer coordination successes seem to move us away from the artificiality of the laboratory. A corollary to that statement is that, to some extent, the initial results of coordination experiments, while extraordinarily successful in generating discussion and subsequent research, were too hastily declared proof positive of the pervasiveness of coordination failure. The literature that has emerged during approximately the past 15 years seems to suggest, in our view, the opposite. Of course, it was the initial results that prompted the follow-up work that helped us understand the determinants of (laboratory) coordination successes, and failures, much better than we did then. Of note in this context is the contribution of the experimental method, which allows us to design and implement experiments that are clean, in that we can study cause and effect in ways that are rarely achievable in real life. Yes, the experimental method has its problems (its lack of external validity being an important one). In principle, however, whatever objection one has against a particular experiment can be evaluated experimentally (e.g., using more realistic subject pools or stimuli).

Future research on coordination games, in our view, would benefit from the application of MouseLab or eye-tracking technologies as well as from gathering sociodemographic and cognitive ability data (e.g., sex, age, income, income proxies, and short-term memory capacity) that could identify correlations with behavior that have been identified for other types of games (for two noteworthy attempts, see Cooper, 2006; Dufwenberg & Gneezy, 2005). For example, recent studies (Devetag, Di Guida, & Polonio, 2013) have attempted to open the “black box” of decision processes leading to focal point selection by means of eye-tracking technologies that enable us to infer the type of heuristic reasoning that subjects employed and to rule out specific explanations of observed behavior (e.g., recognition of iterated dominance). More research of this kind would allow us to pinpoint the features that render an outcome “focal” (payoff symmetry, position, etc.) and the type of cognitive process leading to its selection. Not surprisingly, individual decisions are characterized by much heterogeneity, a theme that has gained some currency also for coordination games (e.g., Sakovics & Steiner, 2012; Semeshenko et al., 2010) and is likely to be of considerable interest.

Research on coordination problems would also, in our view, benefit from additional efforts to move laboratory scenarios closer to the real world. For example, it is quite desirable to understand what parameterizations of the coordination games we have described in this chapter are the most suitable to reflect coordination phenomena in real-life markets and other social contexts. In other words, it seems particularly desirable to understand what would be,

for a certain problem, an appropriate parameterization (“calibration”) of the game that tries to model the situation. Every experimental test requires numerous design and implementation decisions, and some of them involve more “realism” than others (Devetag & Ortmann, (2007)). Apart from communication options, the issue of chosen effort remains important. As discussed previously, the choice of an action in a minimum effort game, for example, is typically hypothetical, although it might have tangible consequences. In this sense, effort is “chosen” and not real. As also discussed, Bortolotti et al. (2013) implemented a testbed specifically designed to test coordination problems with real effort as opposed to the traditional lab experiments using chosen effort. The results indicate that outcomes may be sensitive to this specific implementation detail, and that coordination problems in the real world may be far less widespread than is currently thought. More experiments on coordination games with real effort (and distinguishing between cognitive and physical effort) would increase the external validity of our knowledge about coordination games. It is both desirable and predictable that clever researchers will soon conduct more natural field experiments (Harrison & List, 2004), quite possibly matched with complementary laboratory experiments, on coordination phenomena (e.g., within organizations) that could bring us closer to a realistic assessment of the way people solve coordination problems in real-life settings. The promise of such an approach has been, in our view, remarkably well demonstrated (List, 2006), albeit for another class of games.

## ACKNOWLEDGMENTS

We thank Jane Sell for her constructive comments and Julie Ann VanDusky for improving the readability of an earlier version of the manuscript.

## REFERENCES

- Bangun, L., Chaudhuri, A., Prak, P., & Zhou, C. (2006). Common and almost common knowledge of credible assignments in a coordination game. *Economics Bulletin*, 3, 1–10.
- Bardsley, N., Mehta, J., Starmer, C., & Sugden, R. (2010). Explaining focal points: Cognitive hierarchy theory versus team reasoning. *Economic Journal*, 120, 40–79.
- Basteck, C., Daniels, T. R., & Heinemann, F. (2013). Characterising equilibrium selection in global games with strategic complementarities. *Journal of Economic Theory*, 148, 2620–2637.
- Battalio, R. C., Samuelson, L., & Van Huyck, J. (2001). Optimization incentives and coordination failure in laboratory stag hunt games. *Econometrica*, 69, 749–764.
- Berninghaus, S. K., & Ehrhart, K.-M. (1998). Time horizon and equilibrium selection in tacit coordination games: Experimental results. *Journal of Economic Behavior and Organization*, 37, 231–248.
- Berninghaus, S. K., & Ehrhart, K.-M. (2001). Coordination and information: Recent experimental evidence. *Economics Letters*, 73, 345–351.
- Blume, A., & Ortmann, A. (2007). The effects of costless pre-play communication: Experimental evidence from games with Pareto-ranked equilibria. *Journal of Economic Theory*, 132, 274–290.

- Bortolotti, S., Devetag, G., & Ortmann, A. (2013). Group incentives or individual incentives? A real-effort weak-link experiment. *Working Paper*,
- Brandenburger, A. M., & Nalebuff, J. N. (1997). *Co-opetition*. New York: Doubleday.
- Brandts, J., & Cooper, D. J. (2006a). A change would do you good.... An experimental study on how to overcome coordination failure in organizations. *American Economic Review*, 96, 669–693.
- Brandts, J., & Cooper, D. J. (2006b). Observability and overcoming coordination failure in organizations. *Experimental Economics*, 9, 407–423.
- Cachon, G. P., & Camerer, C. F. (1996). Loss-avoidance and forward induction in experimental coordination games. *Quarterly Journal of Economics*, 111, 165–194.
- Camerer, C. (2003). *Behavioral game theory: Experiments in strategic interaction*. Princeton, NJ: Princeton University Press.
- Camerer, C., & Knez, M. (1996). Coordination, organizational boundaries and fads in business practice. *Industrial and Corporate Change*, 5, 89–112.
- Camerer, C., & Knez, M. (1997). Coordination in organizations: A game-theoretic perspective. In Z. Shapira (Ed.), *Organizational decision making* (pp. 158–188). Cambridge, UK: Cambridge University Press.
- Camerer, C., & Lovallo, D. (1999). Overconfidence and excess entry: An experimental approach. *American Economic Review*, 89, 306–318.
- Carlsson, H., & van Damme, E. (1993). Global games and equilibrium selection. *Econometrica*, 61, 989–1018.
- Chaudhuri, A., Schotter, A., & Sopher, B. (2009). Talking ourselves to efficiency: Coordination in intergenerational minimum effort games with private, almost common and common knowledge of advice. *Economic Journal*, 119, 91–122.
- Clark, K., & Sefton, M. (2001). Repetition and signalling: Experimental evidence from games with efficient equilibria. *Economics Letters*, 70, 357–362.
- Colman, A. M., Pulford, B. D., & Rose, J. (2008). Collective rationality in interactive decisions: Evidence for team reasoning. *Acta Psychologica*, 128, 387–397.
- Cooper, D. J. (2006). Are experienced managers experts at overcoming coordination failure? *Advances in Economic Analysis & Policy*, 6.2, Article 6.
- Cooper, R., De Jong, D., Forsythe, R., & Ross, T. (1989). Communication in the battle of the sexes game. *The Rand Journal of Economics*, 20, 568–587.
- Cooper, R., De Jong, D., Forsythe, R., & Ross, T. (1990). Selection criteria in coordination games: Some experimental results. *American Economic Review*, 80, 218–233.
- Cooper, R., De Jong, D., Forsythe, R., & Ross, T. (1992). Communication in coordination games. *Quarterly Journal of Economics*, 107, 739–771.
- Costa-Gomes, M., Crawford, V., & Broseta, B. (2001). Cognition and behavior in normal form games: An experimental study. *Econometrica*, 69, 1193–1235.
- Devetag, G. (2003). Coordination and information in critical mass games: An experimental study. *Experimental Economics*, 6, 53–73.
- Devetag, G. (2005). Precedent transfer in coordination games: An experiment. *Economics Letters*, 89, 227–232.
- Devetag, G., Di Guida, S., & Polonio, L. (2013, May). *An eye-tracking study of feature-based choice in one-shot games*. LEM working paper, Sant'Anna School of Advanced Studies.
- Devetag, G., & Ortmann, A. (2007). When and why: A critical survey on coordination failure in the laboratory. *Experimental Economics*, 10, 331–344.
- Duffy, J., & Feltovich, N. (2002). Do actions speak louder than words? Observation vs. cheap talk as coordination devices. *Games and Economic Behavior*, 39, 1–27.

- Duffy, J., & Feltovich, N. (2006). Words, deeds and lies: Strategic behavior in games with multiple signals. *Review of Economic Studies*, 73, 669–688.
- Duffy, J., & Ochs, J. (2012). Equilibrium selection in static and dynamic entry games. *Games and Economic Behavior*, 76, 97–116.
- Dufwenberg, M., & Gneezy, U. (2005). Gender and coordination. In A. Rapoport, & R. Zwick (Eds.), In *Experimental business research Vol. 3*. (pp. 253–262). Boston: Kluwer.
- Engelmann, D., & Normann, H. (2010). Maximum effort in the minimum-effort game. *Experimental Economics*, 13, 49–59.
- Farrell, J., & Saloner, G. (1986). Installed base and compatibility: Innovation, product preannouncements and predation. *American Economic Review*, 76, 940–955.
- Gneezy, U., & Rustichini, A. (2000). Pay enough or don't pay at all. *Quarterly Journal of Economics*, 115, 791–811.
- Goeree, J. K., & Holt, C. A. (2005). An experimental study of costly coordination. *Games and Economic Behavior*, 51, 349–364.
- Harrison, G. (2005). Field experiments and control. *Research in Experimental Economics*, 10, 17–50.
- Harrison, G. W., Johnson, E., McInnes, M. M., & Rutström, E. E. (2005). Risk aversion and incentive effects: Comment. *American Economic Review*, 95, 897–901.
- Harrison, G., & List, J. (2004). Field experiments. *Journal of Economic Literature*, 42, 4.
- Harsanyi, J., & Selten, R. (1988). *A general theory of equilibrium selection in games*. Cambridge, MA: MIT Press.
- Heinemann, F. (2012). Understanding financial crises: The contributions of experimental economics. *Annals of Economics and Statistics*, 107–108, 7–29.
- Heinemann, F., Nagel, R., & Ockenfels, P. (2004). The theory of global games on test: Experimental analysis of coordination games with public and private information. *Econometrica*, 72, 1583–1599.
- Hertwig, R., & Ortmann, A. (2001). Experimental practices in economics: A challenge for psychologists? *Behavioral and Brain Sciences*, 24, 383–403.
- Hertwig, R., & Ortmann, A. (2008). Deception in experiments: Revisiting the arguments in its defense. *Ethics and Behavior*, 18, 59–92.
- Holt, C., & Laury, S. (2002). Risk aversion and incentive effects in lottery choices. *American Economic Review*, 92, 1644–1655.
- Holt, C., & Laury, S. (2005). Risk aversion and incentive effects: New data without order effects. *American Economic Review*, 95, 902–912.
- Johnson, E. J., Camerer, C., Sen, S., & Rymon, T. (2002). Detecting failures of backward induction: Monitoring information search in sequential bargaining. *Journal of Economic Theory*, 104, 16–47.
- Kahneman, D. (1988). Experiments in economics: A psychological perspective. In R. Tietz, W. Albers, & R. Selten (Eds.), *Bounded rational behavior in experimental games and markets* (pp. 11–18). New York: Springer.
- Kamecke, U. (1997). Rotation: Matching schemes that efficiently preserve the best response structure of a one-shot game. *International Journal of Game Theory*, 26, 409–417.
- List, J. (2006). The behavioralist meets the market: Measuring social preferences and reputation effects in actual transactions. *Journal of Political Economy*, 114, 1–37.
- Mehta, J., Starmer, C., & Sugden, R. (1994). The nature of salience: An experimental investigation of pure coordination games. *American Economic Review*, 84, 658–673.
- Morris, S., & Shin, H. S. (2003). Global games: Theory and applications. In M. Dewatripont, L. Hansen, & S. Turnovsky (Eds.), *Advances in economics and econometrics (Proceedings*

- of the Eighth World Congress of the Econometric Society).* Cambridge, UK: Cambridge University Press.
- Ochs, J. (1995). Coordination problems. In J. K. Kagel & A. E. Roth (Eds.), *Handbook of experimental economics* (pp. 195–252). Princeton, NJ: Princeton University Press.
- Ortmann, A., & Gigerenzer, G. (1997). Reasoning in economics and psychology: Why social context matter. *Journal of Institutional and Theoretical Economics*, 153, 700–710.
- Ortmann, A., & Hertwig, R. (2002). The costs of deception: Evidence from psychology. *Experimental Economics*, 5, 111–131.
- Rankin, F., Van Huyck, J. B., & Battalio, R. C. (2000). Strategic similarity and emergence of conventions: Evidence from payoff perturbed stag hunt games. *Games and Economic Behavior*, 32, 315–337.
- Rydval, O., & Ortmann, A. (2004). How financial incentives and cognitive abilities affect task performance in laboratory settings: An illustration. *Economics Letters*, 85, 315–320.
- Sakovics, J., & Steiner, J. (2012). Who matters in coordination problems? *American Economic Review*, 102(7), 3439–3461.
- Schelling, T. (1960). *The strategy of conflict*. Cambridge, MA: Harvard University Press.
- Schelling, T. (1978). *Micromotives and macrobehavior*. New York: Norton.
- Schmidt, D., Shupp, R., Walker, J. M., & Ostrom, E. (2003). Playing safe in coordination games: The role of risk dominance, payoff dominance, social history, and reputation. *Games and Economic Behavior*, 42, 281–299.
- Semeshenko, V., Garapin, A., Ruffieux, B., & Gordon, M. (2010). Information-driven coordination: Experimental results with heterogeneous individuals. *Theory and Decision*, 69, 119–142.
- Stango, V. (2004). The economics of standards wars. *Review of Network Economics*, 3, 1–19.
- Van Huyck, J. B., Battalio, R. C., & Beil, R. O. (1990). Tacit coordination games, strategic uncertainty, and coordination failure. *American Economic Review*, 80, 234–248.
- Van Huyck, J. B., Battalio, R. C., & Beil, R. O. (1991). Strategic uncertainty, equilibrium selection, and coordination failure in average opinion games. *Quarterly Journal of Economics*, 106, 885–911.
- Van Huyck, J. B., Battalio, R. C., & Beil, R. O. (1993). Asset markets as an equilibrium selection mechanism: Coordination failure, game form auctions, and tacit communication. *Games and Economic Behavior*, 5, 485–504.
- Van Huyck, J. B., Battalio, R. C., & Rankin, F. W. (2001). *Evidence on learning in coordination games*. Texas A&M University laser script.
- Van Huyck, J. B., Gillette, A., & Battalio, R. C. (1992). Credible assignments in coordination games. *Games and Economic Behavior*, 4, 606–626.
- Weber, R. (2006). Managing growth to achieve efficient coordination in large groups. *American Economic Review*, 96, 114–126.

## Chapter 17

# Experimental Studies of Media Stereotyping Effects

Srividya Ramasubramanian and Chantrey J. Murphy

*Texas A&M University, College Station, Texas*

## I INTRODUCTION

Media psychology is an exciting and challenging area of study today, given the ubiquitous, complex, and dynamic nature of media content. Media are increasingly becoming an integral part of everyday life in a highly networked, global, interactive digital new media world in which meanings and interpretations of media messages vary across situations, audiences, and media formats. The media effects tradition focuses on examining “the social or psychological changes that occur in consumers of media message systems as a result of being exposed to, processing or acting on those mediated messages” (Bryant & Zillmann, 1986, p. 13). These effects could be direct or indirect; short-term or long-term; intended or unintended; immediate or delayed; and behavioral, cognitive, or affective (Bryant & Thompson, 2002). Lab-based experiments are very popular within media effects literature and are seen as the gold standard for drawing causal inferences. One study showed that 29% of media effects studies published from 1993 to 2005 in top communication journals used experimental methods (Potter & Riddle, 2007). Although experiments have been used in many contexts, such as media violence, sexual media, political propaganda, persuasive messages, and health campaigns, the focus of this chapter is on media stereotyping effects.

Along with peers, family, and co-workers, mass media act as key socializing agents in shaping individuals’ attitudes, perceptions, and behaviors. They are an important (and often the primary) source of information for individuals about places, peoples, and cultures, especially when direct contact is absent. Several studies demonstrate that mainstream media has historically marginalized, trivialized, demeaned, and underrepresented minority groups (Dill & Thill, 2007; Dixon & Linz, 2000; Entman & Rojecki, 2001; Mastro & Greenberg, 2000;

(Ramasubramanian, 2005). The idea that media play a significant role in the formation and maintenance of cultural stereotypes and prejudicial feelings toward out-groups has received significant research attention.

Media stereotyping is a subfield of media effects scholarship that examines how stereotypical images and words in mediated messages influence viewers' real-world attitudes and behaviors. This body of literature typically uses a social-cognitive, information-processing approach to examine how media stereotypes affect audiences attitudinally, cognitively, emotionally, and behaviorally. The use of lab-based experiments along with statistical advances in modeling techniques have enhanced the ability of media psychologists to systematically investigate the message features, situational contexts, and audience characteristics that influence cause–effect relationships between stereotypical media content and audience outcomes.

This chapter provides an overview of methodological and theoretical advances in lab-based experimental research on media stereotyping processes. It introduces the reader to some key terms and concepts within this body of literature. The theoretical perspectives used and developed by media stereotyping researchers are elaborated upon in the next section. This is followed by a discussion of the methodological advances that explain some of the standard designs, typical media stimuli, and key dependent variables in experimental work on media stereotyping. The chapter concludes with implications for new digital media contexts and possible future directions.

## II KEY CONCEPTS

A basic concept to be defined and discussed is *stereotype*. Stereotypes have been defined as faulty overgeneralizations (Allport, 1954; Gardner, 1973); cognitive representations that are stored in and retrieved from memory (Dovidio, Evans, & Tyler, 1986; Taylor & Crocker, 1981); specific, concrete exemplars (Bodenhausen, Schwarz, Bless, & Waeke, 1995; Kahneman & Miller, 1986; Linville, Fischer, & Salovey, 1989); and prototypes that are averaged representations of social groups (Fiske & Taylor, 1991; Hamilton & Sherman, 1996; Hilton & von Hippel, 1996). Stereotypes can be positive (e.g., “African-Americans are rhythmic”) or negative (e.g., “Women are irrational”). Recently, media scholars have also started paying attention to *countertereotypes*, which are stereotype-disconfirming information that challenges existing cultural expectations (Aubrey & Harrison, 2004; Bodenhausen et al., 1995; Ramasubramanian, 2007).

Closely related to stereotypes is the concept of *prejudice*. Prejudice involves an affective component and often includes negative evaluations of members of a group. It has been defined as an emotionally rigid predisposition toward groups (McKenzie-Mohr & Zanna, 1990; Rudman & Borgida, 1995; Simpson & Yinger, 1958). Stangor (2000) associates prejudice with negative emotional feelings such as dislike, uneasiness, anger, disgust, and hatred against people based on their group affiliation. Blatant prejudice involves expressing negative

feelings quite candidly and endorsing derogatory statements about stigmatized groups (Kovel, 1970). Here, there is often a perceived threat from, rejection of, and lack of desire for intimate contact with the stereotyped group (Pettigrew & Meertens, 1995). Modern forms of prejudice tend to be more abstract, political, ambivalent, aversive, covert, and subtle (Gaertner & Dovidio, 1986; McConahay, 1986).

Another challenge is defining mass media, given their dynamic, rapidly evolving, and complex nature. Considering the proliferation, personalization, and convergence of media types, our understandings of *media* and *audiences* are constantly being challenged and redefined. Mediated communication, often distinguished from face-to-face communication, can be either interpersonal (e.g., telephone) or meant for mass audiences (e.g., television and radio). In this chapter, we focus primarily on mass media stereotypes and their effects. Mass media types can include print media (e.g., books, magazines, and newspapers), broadcasting media (e.g., television, radio, and films), and new media (e.g., Internet, smartphones, computers, and video game consoles). Media stereotyping research has primarily focused on the effects of newspapers, television programming, and films, although research about other media formats is not unheard of.

### III THEORETICAL ADVANCES

Among media psychologists, the notion that media portrayals about various social groups play an important role in shaping attitudes toward those groups has garnered much attention. Researchers have employed a variety of theoretical mechanisms to explore the assumption that individuals for whom mediated messages serve as the primary source of information about out-groups are likely to apply them in their impressions and judgments about these groups.

#### A Cultivation and Mental Models

Scholarship on *cultivation perspectives* suggests that long-term exposure to media content, especially television, can have cumulative effects that distort social reality perceptions and increase stereotypical beliefs (for a review, see Gerbner, Gross, Morgan, Signorielli, & Shanahan, 2002). Research supporting this hypothesis shows that heavy as compared to light television viewers have more stereotypical perceptions about racial minorities, sexist beliefs about women, and negative attitudes toward those with mental illnesses (Armstrong, Neuendorf, & Brentar, 1992; Busselle & Crandall, 2002; Diefenbach & West, 2007; Gerbner et al., 2002; Ward, 2002). These real-world estimates (*first-order effects*) can also affect beliefs, values, and policy preferences (*second-order effects*).

Although most cultivation research has relied on self-reported surveys, experimental research has played a critical role in explaining the underlying mechanisms. Accessibility-based heuristic activation models suggest that

frequency, relevance, and vividness of stereotypes influence social reality estimates for heavy media users (Busselle & Shrum, 2003; Rothman & Hardin, 1997; Shrum & O'Guinn, 1993). Such models also support *exemplification theory*, which has examined how the nature and number of exemplars influence audience attitudes. Vivid, emotion-laden case reports influence audiences, especially when direct, firsthand experience is lacking (Fujioka, 1999; Gibson & Zillmann, 1994; Zillmann, Gibson, Sundar, & Perkins, 1996). The *mental models approach* is another explanatory mechanism that builds on the idea of *construct accessibility* (Mastro, Behm-Morawitz, & Ortiz, 2007; B. Roskos-Ewoldsen, Davies, & Roskos-Ewoldsen, 2004). *Mental models* are dynamic cognitive representations relating to knowledge about people, objects, issues, and events. Information from a variety of sources such as firsthand experiences and media stereotypes can be synthesized into an individual's mental model in influencing real-world perceptions.

## B Media Priming and Stereotype Activation

Media priming is the most popular theoretical approach to study media stereotyping processes. *Priming* has been defined as the “effect of some preceding stimulus or event on how we react, broadly defined, to some subsequent stimulus” (D. R. Roskos-Ewoldsen, Roskos-Ewoldsen, & Carpentier, 2002, p. 97). Media stimuli act as primes that evoke certain thoughts and feelings, which in turn play a role in coloring impressions formed about issues, objects, or people. Typically, a stereotypical media message serves as the prime, after which participants evaluate an ambiguous situation or judge a target person from the stereotyped group.

According to the *neo-associationistic model* of priming, through repeated associations of related thoughts and feelings, a cognitive-affective network of nodes is formed in viewers' minds (Jo & Berkowitz, 1994). When one node is activated by the media message, other related nodes in the network also get triggered through a process known as *spreading activation*. For instance, Valentino (1999) found that crime news stories primed other racially coded issues such as welfare, which influenced evaluations of political candidates. Stereotypical media exemplars have been shown to also activate affective reactions such as feelings of hostility and pity (Johnson, Bushman, & Dovidio, 2008; Ramasubramanian & Oliver, 2007) apart from *implicit attitudes* (Brown Givens & Monahan, 2005; Ramasubramanian, 2007; Ramasubramanian & Oliver, 2007).

The majority of the research on media priming has focused on how negative stereotypes in the news media prime judgments of African-American targets, especially relating to criminality (Dixon, 2006a, 2006b, 2007; Gilliam & Iyengar, 2000; Johnson, Adams, Hall, & Ashbum, 1997; Johnson et al., 2008; Peffley, Shields, & Williams, 1996). Constant exposure to media stereotypes makes them chronically accessible, especially for heavy media consumers. Even

subtle, implicit cues such as varying the racial composition (African-American or White American) of criminal suspects in the news (Dixon & Azocar, 2007), facial features and skin tone of alleged offenders (Oliver, Jackson, Moses, & Dangerfield, 2004), and racial/ethnic identity of subjects in the photograph of online news stories (Abraham & Appiah, 2006) are sufficient to influence subsequent social reality judgments.

## C Social Identity and Social Cognitive Theory

Social identity theory proposed by Tajfel and Turner (1986) suggests that individuals experience collective identity based on their membership in a group, such as racial/ethnic and gender identities. Social identity leads individuals to categorize themselves and other salient groups into “us” versus “them.” *Self-categorization* based on group membership might be so salient that it can get activated automatically even with subtle stimuli. To maintain positive social identity, people engage in intergroup comparisons that demonstrate a favorable bias toward their in-group, display discriminatory behaviors toward out-groups, and use coping mechanisms such as internal/external causal attributions for group failures (Brewer, 1979; Brewer, Manzi, & Shaw, 1993; Fiske & Taylor, 1991; Hewstone, 1990).

Media stereotyping studies have applied social identity perspectives to understand effects on both majority and minority group members. Group identity is especially salient for members of minority groups, and studies show that they prefer content featuring members of their minority in-groups in the media (Appiah, 2001, 2002; Fujioka, 2005). Audience members from minority groups are conscious of features that might mark them as distinct from the majority group and are particularly sensitive to how they are represented in popular media, in which they are often typically invisible. According to the *ethnolinguistic identity theory*, viewing media programs that feature members of their group increases their in-group vitality, especially when depicted in a positive light (Abrams, Eveland, & Giles, 2003; Giles, Bourhis, & Taylor, 1977).

With regard to research on majority group members, when media representations of out-groups “accommodate” in-group norms, minority group members in real-life are evaluated less stereotypically (Coover, 2001). Media stereotyping serves as an avenue for categorizing other groups, especially when the stereotype serves the in-group positively and the out-group negatively. For example, Mastro (2003) showed that White audiences, especially those with higher racial identification, would have a greater tendency to judge Latinos in a negative light after exposure to televised portrayals of Latino criminality and also reported higher self-esteem when exposed to Latino criminality on television.

Another theoretical perspective that tries to integrate real-world experiences with mediated ones in shaping identities and behaviors is the *social cognitive theory*. It suggests that people cognitively process information and internalize responses to situations based on observations, even when they do

not experience them firsthand, and adapt them to their own contexts (Bandura, 1977, 2002). Although very complex and broad in scope, some key concepts from this theory, such as abstract modeling, inhibitory and disinhibitory processes, vicarious learning, and positive/negative reinforcement, have been applied to media stereotyping studies. For example, Ortiz and Harwood (2007) examined whether positive intergroup interactions role-modeled in the media would lead to positive attitudinal outcomes through abstract modeling and identification. Fujioka (1999) found that the nature of vicarious contact (positive or negative) with African-Americans via television portrayals shaped Japanese international students' attitudes toward this group. Behm-Morawitz and Mastro (2009) found that sexualized portrayals of female characters in video games can negatively influence self-esteem and self-efficacy in female gamers. Ward and Friedman (2006) found that adolescents who viewed stereotypical media portrayals of women as sex objects were more likely to be supportive of sexist behaviors.

## D Counter-Stereotypes, Parasocial Contact, and Stereotype Change

The question of whether counter-stereotypes or atypical exemplars shift attitudes in the positive direction or lead to more modern prejudice has been an important debate within the field (Bodenhausen et al., 1995; Holt, 2013; Ramasubramanian, 2007). Bodenhausen and colleagues found support for the *appraisal generalization* perspective that suggests that activation of positive media exemplars can lead to positive shifts. Ramasubramanian combined a message-centered approach (e.g., the use of counter-stereotypical positive exemplars of stigmatized groups) along with an audience-centered approach (e.g., media literacy training) for effective prejudice reduction at the implicit and explicit level. Combining real-world and media-based strategies, Nathanson Wilson, McGee, and Sebastian (2002) found that when an adult experimenter used active mediation by providing a contradictory message to televised stereotypes, children reported lesser acceptance of gender stereotypes.

Allport (1954), in his intergroup contact theory, explains that interpersonal contact can be one of the most effective ways to reduce prejudice between majority and minority group members. Applying this idea to mediated contact, Riggle, Ellis, and Crawford (1996) found that exposure to positive gay characters through a documentary film had a significant impact on attitudes, even after controlling for prior real-world contact with gay men. Similarly, Mastro and Tropp (2004) found that levels of real-world inter-racial contact influenced White participants' attitudes toward African-Americans after viewing TV sitcoms featuring stereotypical versus nonstereotypical Black characters. Schiappa, Gregg, and Hewes (2005) combined contact hypothesis with parasocial interaction, which emphasizes positive emotional bonds formed with media personalities (Horton & Wohl, 1956) to build the *parasocial contact hypothesis*.

They found that parasocial contact with minority group members through television programs leads to prejudice reduction among majority group members.

## IV METHODOLOGICAL ADVANCES

This section provides an overview of the standard designs, typical stimuli, common dependent variables, sample populations, and procedures used by media psychologists who study media stereotyping processes.

### A Standard Designs and Procedures

Most of the lab-based experiments on media stereotyping use between-group factorial designs, although mixed designs are almost as common. In its simplest form of a  $2 \times 2$  design, the typical factors are stereotypicality of the media message (stereotypical, counter-stereotypical, or neutral) and group affiliation of the target in the message (majority group member and minority group member). Complex experimental designs with many message-related factors and audience-related variables are not uncommon. For example, [Fujioka \(2005\)](#) used a repeated measured mixed experimental design using 2 (arousal: calm or arousing)  $\times$  2 (valence of stories: positive or negative)  $\times$  3 (story type)  $\times$  2 (respondent's identity: Mexican American or White)  $\times$  4 (presentation order).

Almost all experimental studies on media stereotyping use a post-only design, perhaps to avoid social desirability biases associated with sensitizing participants to measures of prejudice and stereotypes. Occasionally, researchers screen their participants prior to the actual experiment to pinpoint preexisting attitudes or predispositions in their sample population. This is done to test whether the experimental manipulations do indeed affect change in preexisting attitudes ([Oliver & Fonash, 2002](#); [Riggle et al., 1996](#); [Schiappa et al., 2005](#)). For example, Oliver and Fonash measured anti-Black attitudes in the first phase of their study, 2 weeks prior to the actual experiment, to ensure that there were no preexisting differences between White and Black participants in terms of their perceptions of stimulus materials.

Most media stereotype experimental outcomes result from short-term effects of experimental treatments. The dependent measure follows the experimental treatment within moments. This is especially true of media priming studies, which by their very definition are concerned with measuring short-term effects. These short-term effects may stem from the interventions/treatments activating preexisting stereotypical attitudes. Participants are typically exposed to the stimulus in a lab-based setting in one single sitting, although some studies do expose participants to media stereotypes over multiple sessions (e.g., see [Schiappa et al., 2005](#)).

### B Sample

The vast majority of sample populations used in this area of research are university undergraduate students, often enrolled in introductory communications

courses using convenience sampling and compensated via course credits. In other research designs, student populations are avoided, and a more general population is sought out to increase the generalizability of the results (Gilliam & Iyengar, 2000; Hurwitz & Peffley, 1997; Kahn, 1994; Valentino, 1999). Such participants are recruited from the local community through newspapers, posters, newsletters, fliers, and other announcements. They are typically compensated monetarily for their participation.

The participants are almost always majority group members. Almost all studies that focus on racial stereotypes, for instance, use White audiences as their sample population. In many cases, if samples are not purposefully restricted to only White participants, participants who identify as White are the largest racial representation, often comprising more than 70% of the sample population. However, a few studies do examine the responses of minority group members such as Black Americans, Mexican Americans, and Latino Americans (Fujioka, 2005; Richeson & Polleydore, 2002; Sanders & Ramasubramanian, 2012). The samples are typically equally distributed across gender categories. When other participant characteristics are reported, they indicate that participants are relatively well educated, heterosexual in sexual orientation, and more Democratic than Republican in political affiliation. Given that prejudicial attitudes are often strongly correlated with political orientation and level of education, the lack of heterogeneity in these samples is a serious threat to ecological validity.

## C Independent Variables

The primary independent variable in almost all these experiments is “exposure to media content.” Typically, the media content is manipulated to be either stereotypical or non-stereotypical. Apart from stereotypical media messages and the neutral control condition, some researchers have also started including a third variation of the media message: counter-stereotypical media content. For example, Ramasubramanian (2007) presents participants with counter-stereotypical news stories on African-Americans and Asian Indians to examine their effects on implicit racial attitudes. Holt (2013) also presents stereotypical, counter-stereotypical, and neutral news stories to participants. Most of the research has focused on racial/ethnic stereotyping, especially about African-Americans, and on gender stereotypes. In contrast, there are relatively fewer studies on media stereotyping effects relating to sexual orientation, age, or social class.

When the group context is varied within the media content, it is typically manipulated as either majority group members or minority group members. For example, the race of the crime suspect (Holt, 2013; Oliver & Fonash, 2002), the name of the alleged suspect in a vignette (Ford, 1997), or the race of an evacuee in a disaster story (Johnson et al., 2008) is manipulated to be Black or White. Similarly, the gender of a gubernatorial candidate is varied to be either male or female (Kahn, 1994). Kilbourne (1990) does not vary the gender of the model

in advertisements but, rather, the role played by the target female exemplar to be either an in-role (housewife) or an out-role (professional).

Another common method for manipulating this independent variable is to vary the composition of groups represented in a media message. Typically, the conditions include majority group members only, minority group members only, or a mix of both. For instance, [Dixon and Azocar \(2007\)](#) varied the racial composition of suspects in news stories to be majority Black or majority White. Very rarely, studies include media portrayals of more than one minority group, which allows for studying comparative stereotyping effects among minority groups ([Ramasubramanian & Oliver, 2007](#)).

Other common participant-related independent variables are level of participants' identification ([Mastro, 2003](#)), level of prejudice ([Mastro & Tropp, 2004](#); [Mendelberg, 1997](#); [Peffley et al., 1996](#)), prior media exposure ([Dixon, 2006b](#); [Dixon & Maddox, 2005](#); [Johnson et al., 2008](#)), frequency of intergroup contact ([Mastro & Tropp, 2004](#)), quality of intergroup contact, and group typicality ([Ortiz & Harwood, 2007](#)). Although participant-related demographics such as participant's gender, age, income, and race serve as control variables for most studies, other attitudinal measures, such as social conservatism, party affiliation, political ideology, media habits, issue salience, and motivation to control prejudice, are also measured ([Dixon, 2006b, 2007](#); [Dixon & Azocar, 2007](#); [Mastro & Tropp, 2004](#); [Mendelberg, 1997](#)).

## D Stimuli

Much attention is given to selecting stimuli based on prior research, gauging their appropriateness through pretests, and manipulating subtle aspects of the media stereotypes. For pretests, a separately drawn sample from the larger population is provided stimulus materials and asked to evaluate how consistent the stimulus is with the experimenter's desired outcome. For instance, with photographs that vary the facial features or the name of the protagonist, experimenters interested in racial stereotypes may ask participants to rank how likely it is that the person depicted belongs to a specific racial group ([Ford, 1997](#); [Mastro, 2003](#)). In other situations, pretest participants are instructed to list the common stereotypical attributes associated with a particular target population, which are then used to identify relevant media stories ([Mastro, 2003](#); [Ramasubramanian, 2007](#); [Ramasubramanian & Oliver, 2007](#)).

Most media stimuli introduce a stereotypical image, news story, or program that works to prime the participants. Stimuli are often borrowed from existing media sources, such as television news segments, advertisements, and photographs. Often, visual aspects are edited or modified to suit the experimental condition. For example, [Abraham and Appiah \(2006\)](#) manipulated news stories to have no photos, Black photo, White photo, or both Black and White photo. Researchers take care to make sure that the story size or video clip length or format is kept identical across conditions by maintaining consistency across story types.

Although television news stories (Dixon, 2006a, 2006b; Dixon & Azocar, 2007; Dixon & Maddox, 2005; Gilliam & Iyengar, 2000; Mendelberg, 1997; Oliver & Fonash, 2002; Peffley et al., 1996; Valentino, 1999) and print news stories (Abraham & Appiah, 2006; Holt, 2013; Johnson et al., 2008; Ramasubramanian, 2007; Ramasubramanian & Oliver, 2007) represent the overwhelming majority of the stimuli used in this field, some researchers have used comedy segments (Ford, 1997; Mastro & Tropp, 2004; Richeson & Pollydore, 2002), reality television (Ortiz & Harwood, 2007; Schiappa et al., 2005), crime drama scripts (Mastro, 2003), celebrity photos (Ramasubramanian, 2011), advertisements (Kilbourne, 1990), and documentary films (Riggle et al., 1996).

It is the norm within this subfield to use “foils” or “fluff” stories to mask the purpose of the research. In priming studies, control conditions provide information that is completely unrelated in terms of content to the experimental treatment. This procedure is followed when the control condition provides no additional priming material that may interfere with or taint the desired neutral effects. However, other experimenters, such as [Mastro and Tropp \(2004\)](#), use nonstereotypic content from the same genre or sometimes from the very same program for the neutral condition.

## E Dependent Measures

Measuring participants’ attitudes, whether preexisting or as a result of experimental treatments, often takes the form of paper-and-pencil questionnaires. Increasingly, however, they are administered as computer-based or online questionnaires. Some questions are bogus for the purposes of concealing the true intention of the study, whereas others directly pertain to the dependent measures. They are often couched as “social beliefs inventory” or “quality of life indicators.” The dependent measures commonly used are cognitive in nature relating to attitudes, judgments, or opinions. These are typically measured using Likert-type measures. Recently, there has been more emphasis on affective measures such as the measurement of prejudicial feelings using feelings thermometers. Physiological measures and behavioral effects, although popular elsewhere in media effects studies, have not been used within media stereotyping studies.

With regard to measures of perceptions, typically semantic differentials or agreement levels on a list of trait dimensions are used to measure stereotypical perceptions ([Kahn, 1994](#); [Kilbourne, 1990](#); [Ramasubramanian & Oliver, 2007](#)). For example, Kahn measured perceptions of candidates’ leadership, compassion, and honesty. They are almost always about out-group stereotypes, although occasionally researchers also measure in-group perceptions such as self-esteem ([Mastro, 2003](#)). These perceptions could also relate to the evaluation of the perceived stereotypicality of a media exemplar ([Bodenhausen et al., 1995](#); [Mastro & Tropp, 2004](#)), beliefs such as mean world syndrome ([Dixon, 2006b](#)), causes for crime and failures ([Holt, 2013](#); [Ramasubramanian, 2011](#)), attitudinal measures such as participants’ level of agreement on controversial statements

(Bodenhausen et al., 1995), affective responses such as social distance (Ortiz & Harwood, 2007; Schiappa et al., 2005), and explicit stereotype endorsement (Dalisay & Tan, 2009; Gilliam & Iyengar, 2000; Johnson et al., 2008).

Dependent measures could also be implicit measures such as word-fragment tasks, lexical decision-making, and the Implicit Association Test (IAT) (Brown Givens & Monahan, 2005; Ramasubramanian, 2007). The assumption behind the use of implicit measures is that automatic stereotype activation is beyond the realm of conscious control and cannot be measured effectively using traditional paper-and-pencil self-reported measures. Instead, experimenters use the speed of reaction to stimuli as an indicator of the strength of association between media stimuli and stereotypes that are activated in memory. For instance, Ramasubramanian used a lexical decision task to measure the reaction time of participants to stereotypical words. Immediately following the media priming task, participants were instructed to distinguish words from non-words presented in quick succession on a computer screen. Those exposed to stereotypical media primes in the experimental condition were expected to recognize stereotypical words at a faster rate than those in the control condition.

Cognitive measures could be about judgments about a target person after a priming task. Typically, these items measure the culpability of an alleged suspect after exposure to the stereotypical primes relating to criminality (Dixon, 2006a; Ford, 1997). They could also be the recommended prison sentence for the suspect (Dixon & Maddox, 2005; Peffley et al., 1996); the justifiability of a behavior of a media character in a program (Mastro, 2003); likely future criminal behavior (Peffley et al., 1996); or the expected performance ratings for political candidates on specific issues such as welfare, crime, and taxation (Valentino, 1999).

Another type of cognitive measure is support for various public policies such as endorsement of affirmative action (Dalisay & Tan, 2009; Ramasubramanian, 2011), the death penalty (Dixon, 2006a, 2006b; Dixon & Azocar, 2007), foreign relations (Mendelberg, 1997), welfare (Mendelberg, 1997), and crime (Gilliam & Iyengar, 2000; Holt, 2013). Relatedly, such items could also measure support for harmful treatment of people in need (Johnson et al., 2008) and causal attributions for out-group failures (Knobloch-Westerwick & Taylor, 2008; Ramasubramanian, 2011).

Some scholars have also measured memory by asking participants to indicate whether or not a particular criminal suspect appeared in the news stories (Oliver & Fonash, 2002) or the racial likelihood of a suspect as Black or White (Dixon, 2007). A very interesting variation of this study asks participants to reconstruct the facial features of the crime suspect who they read about and then measure the extent to which they are Afro-centric (Oliver et al., 2004).

Affective measures could include discrete emotions such as hate, contempt, fear, pity, anxiety, or discomfort (Ortiz & Harwood, 2007; Peffley et al., 1996; Ramasubramanian & Oliver, 2007; Richeson & Pollydore, 2002). For example, Ramasubramanian and Oliver measure benevolent prejudice that includes sympathy, guilt, and pity toward the stigmatized groups. Feelings thermometers have

been used by scholars to measure the level of warmth toward various groups in society (Mendelberg, 1997; Ramasubramanian & Oliver, 2007). Affective reactions could also be more indirect measures such as respondents' emotional concern about the story (Dixon & Maddox, 2005) or about an issue such as crime (Holt, 2013).

## V IMPLICATIONS FOR NEW MEDIA AND FUTURE DIRECTIONS

Much of the existing experimental work on media stereotyping has been conducted in the context of "traditional" media formats. However, no longer are audiences restricted to "old" media resources such as network and cable television, print media such as magazines and newspapers, and radio. They now have access to new media such as blog spaces, mobile phones, social networks such as Facebook and Twitter, video games, as well as Internet video streaming sites such as Netflix, YouTube, or Hulu. The advancement of interactive, digital, global, and personalized media presents new opportunities for scholars to test whether existing media stereotyping theories hold true in these contexts and also to use these formats for methodological advances in experimental research.

In extensive reviews, Metzger (2009) and Chaffee and Metzger (2001) highlight what media scholars posit are difficulties surrounding the theoretical and methodological concerns of new media. Theoretically, the greater availability and dissemination of information across new media contexts, mediums, populations, and points in time is likely to bring about challenges to current areas of media effects studies. In an increasingly user-driven new media context, audience members are more selective and less likely to grow up in a commonly shared symbolic environment that cultivates stereotype formation. However, a counterpoint to this new media concern follows that cultivation is still likely to take place; instead of an aggregate idea of how groups of people interact, the audience is able to tailor its media in such a way as to cultivate its own understanding of how groups engage with one another.

The media "audience" member has been conceptualized as being restricted to what is presented by media outlets and to take this information at face value. Alternatively, media "users" in the current media contexts are interactive individuals who exercise greater authority when deciding what type of messaging they will allow themselves to be exposed to. The interactivity of media lends itself to greater audience agency so that audiences are able to alter the media content in which they are engaging. Users are able to mitigate and manipulate the information provided via media outlets while also having the ability to broadcast information to other audiences via indirect (e.g., Facebook newsfeeds) or direct (e.g., blog postings) means. However, mere selectivity and interactivity might be different from exercising media literacy skills that critically examine the messages that they are presented with, their meanings, and their purposes.

While theoretical implications of new media contexts develop, methodological considerations focus on developing research designs that accommodate the various ways that audiences interact with media content. To examine media stereotypes and their effects in such new media as social media sites, chat rooms, video games, and virtual worlds such as Second Life, experimentalists have to become familiar with manipulating stimuli materials and setting up treatment conditions in these contexts. Media scholars are also increasingly recruiting samples using social media such as Facebook and administering their experiments using online, computer-based tools such as Medialab, Survey Monkey, Qualtrics, and Mechanical Turk. Often, these tools are administered in lab-based contexts but also sent as links to participants to be completed on their laptops and mobile devices. Given that there are variations in experimental settings, it is important for researchers to control for when and how experiments are administered to participants.

The shifting media cultural scenario means that stereotypes are much more subtle and covert than ever before. With the emergence of the post-feminist and post-racial era, media scholars are particularly interested in studying the effects of media stereotypes on symbolic, ambivalent, and modern prejudices. In line with this shift in media representations, computer-based tools such as the IAT and lexical decision tasks offer new avenues to go beyond traditional paper-and-pencil measures of explicit stereotype endorsement. They allow researchers to measure implicit, subtle, and aversive forms of prejudices rather than old-fashioned blatant prejudice. These tools capture automatic stereotypes that operate without participants' conscious awareness, even though they might be malleable to situational cues such as counter-stereotypical media exemplars (Banaji & Hardin, 1996; Blair, 2002; Devine, 1989; Ramasubramanian, 2007).

Media stereotyping research tends to largely focus on racial/ethnic stereotypes within U.S. contexts. Future research should examine these topics in other cultural contexts and also on other group affiliations, such as sexual orientation, body size, and social class. Experimenters should go beyond student samples to more generalizable audience populations to increase the ecological validity of their findings. In addition to examining the detrimental effects of media stereotypes on majority and minority group members, researchers should also focus on understanding the role of positive counter-stereotypical media exemplars and media literacy strategies for developing media-based strategies for prejudice reduction.

Including more complex designs such as pre–post or Solomon four-group design would allow for greater control and more confidence in findings. Researchers should explore longitudinal designs that allow for multiple exposures to media content to track dynamic audience reactions over time. Through the use of sophisticated tools such as structural equation modeling, media scholars interested in cause–effect inferences should focus more on building causal chains of relationships to explain the key variables that moderate and mediate the relationship between media stereotypes and audience attitudes.

Future research should focus on the length of effects to explain the conditions under which effects last only for a few minutes versus those that last for much longer.

In conclusion, evolving epistemological notions of media, different conceptualizations of media effects (e.g., the cultivation hypothesis), and the development of sophisticated statistical procedures have the potential to lead to significant theoretical and methodological advances in experimental research on media stereotyping. New models, vocabularies, and theories of media effects research are being developed to theorize interactive, transnational, and new digital media experiences. Research on media stereotyping is shifting in its emphasis from *whether* media stereotypes exist to *how* they influence audiences. In other words, there is a move from media stereotyping *effects* to *processes*. For all these reasons, the future holds much promise and excitement in terms of experimental research on media stereotyping processes.

## REFERENCES

- Abraham, L., & Appiah, O. (2006). Framing news stories: The role of visual imagery in priming racial stereotypes. *Howard Journal of Communications*, 17(3), 183–203.
- Abrams, J. R., Eveland, W. P. J., & Giles, H. (2003). The effects of television on group vitality: Can television empower nondominant groups? In P. J. Kalbfleisch (Ed.), *Communication yearbook*: 27. (pp. 193–219). Mahwah, NJ: Lawrence Erlbaum.
- Allport, G. W. (1954). *The nature of prejudice*. Cambridge, MA: Perseus Books.
- Appiah, O. (2001). Ethnic identification on adolescents' evaluations of advertisements. *Journal of Advertising Research*, 41(5), 7–22.
- Appiah, O. (2002). Black and white viewers' perception and recall of occupational characters. *International Communication Association*, 52(4), 776–793.
- Armstrong, G. B., Neuendorf, K. A., & Brentar, J. E. (1992). TV entertainment, news, and racial perceptions of college students. *Journal of Communication*, 42(3), 153–176.
- Aubrey, J. S., & Harrison, K. (2004). The gender-role content of children's favorite television programs and its links to their gender-related perceptions. *Media Psychology*, 6, 111–146.
- Banaji, M. R., & Hardin, C. D. (1996). Automatic stereotyping. *Psychological Science*, 7(3), 136–141.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84(2), 191–215.
- Bandura, A. (2002). Social cognitive theory in cultural context. *Applied Psychology*, 51(2), 269–290.
- Behm-Morawitz, E., & Mastro, D. (2009). The effects of the sexualization of female video game characters on gender stereotyping and female self-concept. *Sex Roles*, 61(11–12), 808–823.
- Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review*, 6(3), 242–261.
- Bodenhausen, G. V., Schwarz, N., Bless, H., & Wanke, M. (1995). Effects of atypical exemplars on racial beliefs: Enlightened racism or generalized appraisals? *Journal of Experimental Social Psychology*, 31, 48–63.
- Brewer, M. B. (1979). In-group bias in the minimal intergroup situation: A cognitive-motivational analysis. *Psychological Bulletin*, 86(2), 307–324.
- Brewer, M. B., Manzi, J. M., & Shaw, J. S. (1993). In-group identification as a function of depersonalization, distinctiveness, and status. *Psychological Science*, 4(2), 88–92.

- Brown Givens, S. M., & Monahan, J. L. (2005). Priming mammies, jezebels, and other controlling images: An examination of the influence of mediated stereotypes on perceptions of an African American woman. *Media Psychology*, 7, 87–106.
- Bryant, J. & Thompson, S. (Eds.), (2002). *Fundamentals of media effects*. Boston: McGraw-Hill.
- Bryant, J. & Zillmann, D. (Eds.), (1986). *Perspectives on media effects*. Hillsdale, NJ: Erlbaum.
- Busselle, R. W., & Crandall, H. (2002). Television viewing and perceptions about race differences in socioeconomic success. *Journal of Broadcasting & Electronic Media*, 46(2), 265–282.
- Busselle, R. W., & Shrum, L. J. (2003). Media exposure and exemplar accessibility. *Media Psychology*, 5(3), 255–282.
- Chaffee, S. H., & Metzger, M. J. (2001). The end of mass communication? *Mass Communication and Society*, 4(4), 365–379.
- Coover, G. E. (2001). Television and social identity: Race representation as “white” accommodation. *Journal of Broadcasting & Electronic Media*, 45(3), 413–431.
- Dalisay, F., & Tan, A. (2009). Assimilation and contrast effects in the priming of Asian American and African American stereotypes through TV exposure. *Journalism & Mass Communication Quarterly*, 86(1), 7–22.
- Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, 56(1), 5.
- Diefenbach, D. L., & West, M. D. (2007). Television and attitudes toward mental health issues: Cultivation analysis and the third-person effect. *Journal of Community Psychology*, 35(2), 181–195.
- Dill, K. E., & Thill, K. P. (2007). Video game characters and the socialization of gender roles: Young people’s perceptions mirror sexist media depictions. *Sex Roles*, 57(11–12), 851–864.
- Dixon, T. L. (2006a). Psychological reactions to crime news portrayals of Black criminals: Understanding the moderating roles of prior news viewing and stereotype endorsement. *Communication Monographs*, 73(2), 162–187.
- Dixon, T. L. (2006b). Schemas as average conceptions: Skin tone, television news exposure, and culpability judgments. *Journalism & Mass Communication Quarterly*, 83(1), 131–149.
- Dixon, T. L. (2007). Black criminals and white officers: The effects of racially misrepresenting law breakers and law defenders on television news. *Media Psychology*, 10(2), 270–291.
- Dixon, T. L., & Azocar, C. L. (2007). Priming crime and activating blackness: Understanding the psychological impact of the overrepresentation of Blacks as lawbreakers on television news. *Journal of Communication*, 57(2), 229–253.
- Dixon, T. L., & Linz, D. (2000). Overrepresentation and underrepresentation of African Americans and Latinos as lawbreakers on television news. *Journal of Communication*, 50(2), 131–154.
- Dixon, T. L., & Maddox, K. B. (2005). Skin tone, crime news, and social reality judgments: Priming the stereotype of the dark and dangerous Black criminal. *Journal of Applied Social Psychology*, 35(8), 1555–1570.
- Dovidio, J. F., Evans, N., & Tyler, R. B. (1986). Racial stereotypes: The contents of their cognitive representations. *Journal of Experimental Social Psychology*, 22, 22–37.
- Entman, R. M., & Rojecki, A. (2001). *The black image in the white mind: Media and race in America*. Chicago: University of Chicago Press.
- Fiske, S. T., & Taylor, S. E. (1991). *Social cognition* (2nd ed.). New York: McGraw-Hill.
- Ford, T. E. (1997). Effects of stereotypical television portrayals of African-Americans on person perception. *Social Psychology Quarterly*, 60(3), 266–275.
- Fujioka, Y. (1999). Television portrayals and African-American stereotypes: Examination of television effects when direct contact is lacking. *Journalism & Mass Communication Quarterly*, 76(1), 52–75.

- Fujioka, Y. (2005). Emotional TV viewing and minority audience: How Mexican Americans process and evaluate TV news about in-group members. *Communication Research*, 32(5), 566–593.
- Gaertner, S. L., & Dovidio, J. F. (1986). *The aversive form of racism*. San Diego: Academic Press.
- Gardner, R. C. (1973). Ethnic stereotypes: The traditional approach, a new look. *The Canadian Psychologist*, 14(2), 133–148.
- Gerbner, G., Gross, L., Morgan, M., Signorielli, N., & Shanahan, J. (2002). Growing up with television: Cultivation processes. In J. Bryant & D. Zillmann (Eds.), *Media effects: Advances in theory and research*. (pp. 43–68) (2nd ed.). Mahwah, NJ: Erlbaum.
- Gibson, R., & Zillmann, D. (1994). Exaggerated versus representative exemplification in news reports: Perception of issues and personal consequences. *Communication Research*, 21(5), 603–624.
- Giles, H., Bourhis, R. Y., & Taylor, D. M. (1977). Towards a theory of language in ethnic group relations. In H. Giles (Ed.), *Language, ethnicity and intergroup relations* (pp. 307–348). London: Academic Press.
- Gilliam, F. D. J. (1999). The “welfare queen” experiment: How viewers react to images of African-American mothers on welfare. *Nieman Foundation for Journalism at Harvard University*, 53(2), 1–6.
- Gilliam, F. D. J., & Iyengar, S. (2000). Prime suspects: The influence of local television news on the viewing public. *American Journal of Political Science*, 44(3), 560–573.
- Hamilton, D. L., & Sherman, S. J. (1996). Perceiving persons and groups. *Psychological Review*, 103(2), 336.
- Hewstone, M. (1990). The “ultimate attribution error”? A review of the literature on intergroup causal attribution. *European Journal of Social Psychology*, 20, 311–335.
- Hilton, J. L., & Von Hippel, W. (1996). Stereotypes. *Annual Review of Psychology*, 47(1), 237–271.
- Holt, L. F. (2013). Writing the wrong: Can counter-stereotypes offset negative media messages about African Americans? *Journalism & Mass Communication Quarterly*, 90(1), 108–125.
- Horton, D., & Wohl, R. (1956). Mass communication and parasocial interaction: Observations on intimacy at a distance. *Psychiatry*, 19, 215–229.
- Hurwitz, J., & Peffley, M. (1997). Public perceptions of race and crime: The role of racial stereotypes. *American Journal of Political Science*, 41(2), 375–401.
- Jo, E., & Berkowitz, L. (1994). A priming effect analysis of media influences: An update. In B. Jennings & D. Zillman (Eds.), *Media effects: Advances in theory and research (LEA's communication series)* (pp. 43–60). Hillsdale, NJ: Erlbaum.
- Johnson, J. D., Adams, M. S., Hall, W., & Ashburn, L. (1997). Race, media, and violence: Differential racial effects of exposure to violent news stories. *Basic and Applied Social Psychology*, 19(1), 81–90.
- Johnson, J. D., Bushman, B. J., & Dovidio, J. F. (2008). Support for harmful treatment and reduction of empathy toward Blacks: “Remnants” of stereotype activation involving Hurricane Katrina and “Lil’ Kim.” *Journal of Experimental Social Psychology*, 44(6), 1506–1513.
- Kahn, K. F. (1994). Does gender make a difference? An experimental examination of sex stereotypes and press patterns in statewide campaigns. *American Journal of Political Science*, 38(1), 162–195.
- Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review*, 93(2), 136.
- Kilbourne, W. E. (1990). Female stereotyping in advertising: An experiment on male–female perceptions of leadership. *Journalism Quarterly*, 67(1), 25–31.
- Knobloch-Westerwick, S., & Taylor, L. D. (2008). The blame game: Elements of causal attribution and its impact on siding with agents in the news. *Communication Research*, 35(6), 723–744.

- Kovel, J. (1970). *White racism: A psychohistory*. New York: Pantheon.
- Linville, P. W., Fischer, G. W., & Salovey, P. (1989). Perceived distributions of the characteristics of in-group and out-group members: Empirical evidence and a computer simulation. *Journal of Personality and Social Psychology*, 57(2), 165.
- Mastro, D. E. (2003). A social identity approach to understanding the impact of television messages. *Communication Monographs*, 70(2), 98–113.
- Mastro, D. E., Behm-Morawitz, E., & Ortiz, M. (2007). The cultivation of social perceptions of Latinos: A mental models approach. *Media Psychology*, 9, 347–365.
- Mastro, D. E., & Greenberg, B. S. (2000). The portrayal of racial minorities on prime time television. *Journal of Broadcasting & Electronic Media*, 44(4), 690–703.
- Mastro, D. E., & Tropp, L. R. (2004). The effects of interracial contact, attitudes, and stereotypical portrayals on evaluations of Black television sitcom characters. *Communication Research Reports*, 21(2), 119–129.
- McConahay, J. B. (1986). Modern racism, ambivalence, and the modern racism scale. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 91–125). San Diego: Academic Press.
- McKenzie-Mohr, D., & Zanna, M. P. (1990). Treating women as sexual objects: Look to the (gender schematic) male who has viewed pornography. *Personality and Social Psychology Bulletin*, 16(2), 296–308.
- Mendelberg, T. (1997). Executing Hortons: Racial crime in the 1988 presidential campaign. *Public Opinion Quarterly*, 61, 134–157.
- Metzger, M. J. (2009). The study of media effects in the era of Internet communication. In R. L. Nabi & M. B. Oliver (Eds.), *Sage handbook of media processes and effects 2* (pp. 561–576). Thousand Oaks, CA: Sage.
- Nathanson, A. I., Wilson, B. J., McGee, J., & Sebastian, M. (2002). Counteracting the effects of female stereotypes on television via active mediation. *International Communication Association*, 52, 922–937.
- Oliver, M. B., & Fonash, D. (2002). Race and crime in the news: Whites' identification and misidentification of violent and nonviolent criminal suspects. *Media Psychology*, 4(2), 137–156.
- Oliver, M. B., Jackson, R. L., 2nd, Moses, N. N., & Dangerfield, C. L. (2004). The face of crime: Viewers' memory of race-related facial features of individuals pictured in the news. *International Communication Association*, 54, 88–104.
- Ortiz, M., & Harwood, J. (2007). A social cognitive theory approach to the effects of mediated intergroup contact on intergroup attitudes. *Journal of Broadcasting & Electronic Media*, 51(4), 615–632.
- Peffley, M., Shields, T., & Williams, B. (1996). The intersection of race and crime in television news stories: An experimental study. *Political Communication*, 13, 309–327.
- Pettigrew, T. F., & Meertens, R. W. (1995). Subtle and blatant prejudice in Western Europe. *European Journal of Social Psychology*, 25, 57–75.
- Potter, W. J., & Riddle, K. (2007). A content analysis of the media effects literature. *Journalism & Mass Communication Quarterly*, 84(1), 90–104.
- Ramasubramanian, S. (2005). A content analysis of the portrayal of India in films produced in the West. *The Howard Journal of Communications*, 16(4), 243–265.
- Ramasubramanian, S. (2007). Media-based strategies to reduce racial stereotypes activated by news stories. *Journalism & Mass Communication Quarterly*, 84(2), 249–264.
- Ramasubramanian, S. (2011). The impact of stereotypical versus counterstereotypical media exemplars on racial attitudes, causal attributions, and support for affirmative action. *Communication Research*, 38(4), 497–516.

- Ramasubramanian, S., & Oliver, M. B. (2007). Activating and suppressing hostile and benevolent racism: Evidence for comparative media stereotyping. *Media Psychology*, 9(3), 623–646.
- Richeson, J. A., & Polleydore, C.-A. (2002). Affective reactions of African American students to stereotypical and counterstereotypical images of Blacks in the media. *Journal of Black Psychology*, 28(3), 261–275.
- Riggle, E. D. B., Ellis, A. L., & Crawford, A. M. (1996). The impact of “media contact” on attitudes toward gay men. *Journal of Homosexuality*, 31(3), 55–69.
- Roskos-Ewoldsen, B., Davies, J., & Roskos-Ewoldsen, D. R. (2004). Implications of the mental models approach for cultivation theory. *Communications*, 29(3), 345–364.
- Roskos-Ewoldsen, D. R., Roskos-Ewoldsen, B., & Carpentier, F. R. D. (2002). Media priming: A synthesis. In J. Bryant & D. Zillmann (Eds.), *Media effects: Advances in theory and research*. (pp. 97–120) (2nd ed.). Mahwah, NJ: Erlbaum.
- Rothman, A. J., & Hardin, C. D. (1997). Differential use of the availability heuristic in social judgment. *Personality and Social Psychology Bulletin*, 23(2), 123–138.
- Rudman, L. A., & Borgida, E. (1995). The afterglow of construct accessibility: The behavioral consequences of priming men to view women as sexual objects. *Journal of Experimental Social Psychology*, 31, 493–517.
- Sanders, M. S., & Ramasubramanian, S. (2012). An examination of African Americans’ stereotyped perceptions of fictional media characters. *Howard Journal of Communications*, 23(1), 17–39.
- Schiappa, E., Gregg, P. B., & Hewes, D. E. (2005). The parasocial contact hypothesis. *Communication Monographs*, 72(1), 92–115.
- Shrum, L. J., & O’Guinn, T. C. (1993). Processes and effects in the construction of social reality: Construct accessibility as an explanatory variable. *Communication Research*, 20(3), 436–471.
- Shuttleworth, F. K., & May, M. A. (1933). *The social conduct and attitudes of movie fans*. New York: Macmillan.
- Simpson, G. E. & Yinger, J. M. (Eds.), (1958). *Racial and cultural minorities: An analysis of prejudice and discrimination*. New York: Harper.
- Stangor, C. (Ed.), (2000). *Stereotypes and prejudice: Essential readings*. Philadelphia: Psychology Press.
- Tajfel, H., & Turner, J. C. (1986). The social identity theory of intergroup behavior. In S. Worchel & W. G. Austin (Eds.), *Psychology of intergroup relations* (pp. 7–24). Chicago: Nelson-Hall.
- Taylor, S. E., & Crocker, J. (1981). Schematic bases of social information processing. *Social cognition: The Ontario Symposium*: Vol. 1. (pp. 89–134). Hillsdale, NJ: Erlbaum.
- Valentino, N. A. (1999). Crime news and the priming of racial attitudes during evaluations of the president. *Public Opinion Quarterly*, 63, 293–320.
- Ward, L. M. (2002). Does television exposure affect emerging adults’ attitudes and assumptions about sexual relationships? Correlational and experimental confirmation. *Journal of Youth and Adolescence*, 31(1), 1–15.
- Ward, L. M., & Friedman, K. (2006). Using TV as a guide: Associations between television viewing and adolescents’ sexual attitudes and behavior. *Journal of Research on Adolescence*, 16(1), 133–156.
- Zillmann, D., Gibson, R., Sundar, S. S., & Perkins, J. W. J. (1996). Effects of exemplification in news reports on the perception of social issues. *Journalism & Mass Communication Quarterly*, 73(2), 427–445.

## Chapter 18

# Judgment and Decision-Making

Michael K. Lindell

*Texas A&M University, College Station, Texas*

## I INTRODUCTION

All human behavior involves the choice of one object or action instead of another, so much social science research addresses decision-making in some form or other. This problem can be defined as a narrow versus broad scope of judgment and decision-making (JDM) research (Van Boven, Traverses, Westfall, & McClelland, 2013), with the small scope focusing on a specialized area of cognitive psychology (see, e.g., Matlin, 2009) that was initially reviewed by Edwards (1961) and most recently reviewed by Weber and Johnson (2009). However, many JDM topics are also addressed in studies of attention, memory, social psychology, health psychology, industrial–organizational psychology, economics, human factors, engineering, organizational management, accounting, medicine, and disaster studies—which is only a partial list of the broad scope of JDM studies. Nonetheless, this chapter focuses on the narrow scope of JDM, which has been developed and tested primarily through laboratory experimentation. Theoretical advancement has been heavily dependent on experimental control to eliminate alternative interpretations for factors affecting decision and judgments. Some of the experiments involve participants faced with hypothetical scenarios and some involve decisions that directly apply to the laboratory context. In addition, field experiments have been employed.

The following sections define decisions and judgments, review the evolution of JDM research, and summarize some of the findings of research on some of the key elements of decisions. Next, the chapter summarizes findings from the most prominent line of research—heuristics and biases; examines the task and cognitive constraints to effective decisions; and reviews two major theoretical issues—the role of emotion in JDM and the possibility of multiple systems underlying JDM processes. Finally, the chapter concludes with a summary of some major findings and unresolved issues.

A *decision* can be defined as “an irrevocable choice of an action that has value-relevant consequences” (Edwards & Fasolo, 2001, p. 582), although Hastie and Pennington (1995) would include conscious deliberation and an

intent to maximize the desirability of the outcome. Moreover, Simon (1955) would highlight situational constraints of uncertainty, complexity, and time, as well as personal constraints such as decision-makers' (DMs') cognitive abilities. In addition, some would include actions that are effortlessly automatic (Andersen, Moskowitz, Blair, & Nosek, 2007) or even habitual (Wood & Neal, 2007). In some theories, decision-making is a multistage process that is initiated by an opportunity or threat that signals the appropriateness of considering a new course of action rather than continuing current activities.

The core elements of a decision are *alternatives* or *options*, the *attributes* or *outcomes* of the alternatives, the *utilities* or *valences* for the outcomes, *uncontrollable events* that cause the alternatives to have more than one outcome, uncertainty about whether the events will occur (*probabilities* or *expectancies* of occurrence), and a rule for combining the probabilities of the events with the values of the outcomes. The alternatives might be given fully developed, found ready-made, modified to meet situational needs, or developed custom made (Mintzberg, Raisinghani, & Théorêt, 1976). In addition, the values for the attributes or the probabilities of outcome occurrence might be static or change over time.

In principle, decision information can be presented in an  $m$  alternatives by  $n$  attributes matrix. When the outcome of a decision depends on an uncertain event that is defined by its probability of occurrence, the possible states of the uncertain event become the columns of the matrix, and the evaluations of the outcomes of a given decision alternative, given the occurrence of an event state, are displayed in the cells of the matrix (Clemen & Reilly, 2004; Cronbach & Gleser, 1957; Raiffa, 1968). For example, Table 18.1 shows the information needed to make a decision to evacuate from a threatening hurricane. A DM can either evacuate or not, and the hurricane will either strike (with probability  $p$ ) or miss (with probability  $1 - p$ ). Cells A and B both incur the cost of an evacuation, and (assuming the evacuation is completed before landfall) neither incurs a loss of life from the hurricane. Cells C and D avoid the cost of an evacuation, but cell D incurs a loss of life if the hurricane strikes. Cell A is a false-positive error

**TABLE 18.1** Decision Table

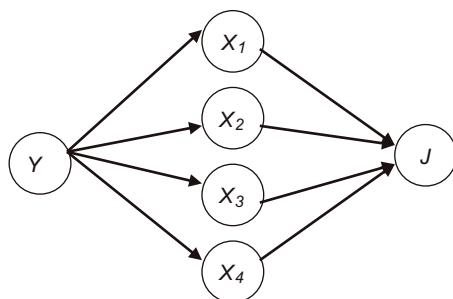
		Hurricane Behavior	
		Miss ( $1 - p$ )	Strike ( $p$ )
Evacuate	Yes	A. Economic cost, No lives lost	B. Economic cost, No lives lost
	No	C. No economic cost, No lives lost	D. No economic cost, Lives lost

because the evacuation incurs an “unnecessary” (in hindsight) cost, whereas cell D is a false-negative error because the failure to evacuate incurs a loss of life. In some cases, decisions such as that depicted in Table 18.1 can incorporate empirical data on costs, losses, and probabilities, but these data must often be obtained from DMs’ judgments.

A *judgment* is an inference from one or more partially reliable cues that provide incomplete and perhaps conflicting information about an unobservable state of nature. This unobservable state could be a DM’s overall preference for a multi-attributed object such as a car defined by fuel economy, styling, and engine performance. Alternatively, it could be an environmental state (e.g., a disease that produces physiological cues) or a future state (e.g., the location of a hurricane’s landfall that can be predicted by environmental cues). Experiments on psychophysical judgments about physical dimensions such as the intensity of light or sound, as well as research on social judgments about phenomena such as the extremity of statements about political or social attitudes, have established that it is very difficult for most DMs to make absolute judgments. Instead, even psychophysical judgments of unidimensional objects are significantly influenced by the DM’s experience and the set of objects being judged. Complex multidimensional social objects are even more influenced by these contextual effects (Eiser & Stroebe, 1972).

One important aspect of judgment is learning from the environment, which can be seen in two major paradigms—Bayesian hypothesis revision and multiple cue probability learning (MCPL). In a prototypical experimental study of Bayesian hypothesis revision (a “bookbag and poker chip” study), DMs are shown two bookbags, one of which has 70% blue and 30% red poker chips, whereas the other has the reverse proportions (Peterson & Beach, 1967). The experimenter selects a bookbag at random, draws a set of chips from it, and asks the DMs to report their judgments of the (posterior) probability of that bookbag being the one that has mostly blue chips. The researcher then compares the DMs’ posterior probabilities to the optimal posterior probabilities computed from Bayes’ theorem. Early research on one-stage single-chain laboratory tasks found that DMs extracted less information than Bayes’ theorem, whereas later research on multistage (cascaded inference) tasks found that DMs provided higher posterior probabilities than Bayes’ theorem (von Winterfeldt & Edwards, 1986). However, recent experiments with subject matter experts making decisions on topics about which they are knowledgeable has found that auditors in cascaded inference tasks extract less information than Bayes’ theorem, which Haynes (2002) attributed to DMs’ greater reliance on source credibility than event diagnosticity.

Another influential thread of judgment research evolved from Brunswik’s (1952) lens model. As indicated in Figure 18.1, an environmental condition ( $Y$ ) produces multiple observable cues ( $X_i$ ) having intercorrelations that might be positive, negative, or zero. The cues also have uncertain relationships with  $Y$ , correlations ( $r_{iy}$ ) known as cue validities, that generally yield imperfect prediction

**FIGURE 18.1** Brunswik's lens model.

( $R^2$  values less than 1.0). In addition, a DM uses the cues to make judgments ( $J$ ), resulting in correlations ( $r_{ij}$ ) known as cue dependencies. To continue the hurricane evacuation example used to illustrate Table 18.1, DMs' judgments of a hurricane's threat to their jurisdiction ( $J$ ) might be based on the storm's strike probability ( $X_1$ ), wind intensity ( $X_2$ ), storm surge height ( $X_3$ ), and current location ( $X_4$ ) (see Lindell & Prater, 2007). To study the DMs' judgmental accuracy in this hurricane threat task, a researcher could ask DMs to judge the hurricane threat from 50 hypothetical hurricanes that vary with respect to these four cues and allow them to observe the correct level of threat after making their judgments, thus receiving outcome feedback. Regression analyses could then be conducted over five blocks of 10 trials each to yield estimates of lens model parameters that characterize changes in the components of judgmental accuracy over successive blocks (for a discussion of the lens model equation, see Stewart, 2001).

## II THE EVOLUTION OF JDM RESEARCH

The concept of rational decision-making has been developed and extended with considerable mathematical sophistication to the problem of a DM choosing among alternative actions whose outcomes are not known with certainty. In its initial form, decision analysis originated with Bernoulli's recommendations for gambling strategy. His recommendation was to choose the alternative that maximizes the quantity  $p_i x_i$ , where  $p_i$  is the probability and  $x_i$  is the value of the  $i$ th outcome. Recognition of certain deficiencies associated with this expected value (EV) principle (for a discussion of JDM paradoxes, see Van Boven et al., 2013) led to the formulation of the expected utility (EU) principle, which recognizes that the value of any outcome to the person may differ from its objective value (Baron, 2008). For example, the difference in utility between \$0 and \$100 is likely to be greater than the difference in utility between \$1,000 and \$1,100. Bernoulli suggested that utility was a logarithmic function of amount.

Although the replacement of (objective) value with (subjective) utility was a significant advance, it had the disadvantage of implicitly assuming that all DMs share a single utility function. This assumption was later addressed by

von Neumann and Morgenstern (1947), who provided a basis for assessing individual differences in utility functions. From a set of six axioms, they developed a system for constructing a utility function from patterns of preferences among a set of gambles.

In judgment research, there was a recognition that probabilities as well as utilities could be taken as subjective and thus differ from one person to another. Those who followed the views of Ramsey (1931) and Savage (1954) considered probability to be the expression of one's degree of belief regarding the occurrence of an event. This viewpoint contrasts with the frequentist view of probability as the limiting value of a frequency distribution. Postulating the existence of meaningful subjective probabilities makes it possible to assess many situations in which frequency data are unavailable. A review of "man as an intuitive statistician" examined subjective assessments of a variety of statistics (e.g., means, variances, and correlations) in a range of situations and concluded that "[i]nferences ... are influenced by appropriate variables and in appropriate directions. But there are systematic discrepancies between normative and intuitive inferences" (Peterson & Beach, 1967, pp. 42–43).

The trend from EV through EU to subjective expected utility (SEU) significantly expanded JDM theory's domain of applicability. Bernoulli's advice applied only to an extremely simplified, repeatable outcome-generating process. In predicting the roll of two dice, a DM knows in advance that the outcome must be between 2 and 12, inclusive. Moreover, once the dice have been rolled, there will be little disagreement among DMs about which outcome occurred. Estimates of the relative likelihood of each outcome can be determined either from logical considerations (one face of each die is "equally likely" to occur) or from observed frequencies of occurrence (the dice can be rolled repeatedly to generate an empirical frequency distribution).

JDM researchers have frequently sought to contrast *normative* (what DMs "should" do) and *descriptive* (what DMs actually do) models. However, judgment research has focused on environmental adaptation; DMs should weight cues according to their validities. By contrast, decision research has focused on internal coherence; DMs should decide consistent with their beliefs and values using an evaluation and combination such as SEU or an alternative such as rank dependent utility (for a discussion, see Baron, 2008, Chapters 10 and 11). This is especially true if DMs need to be very accurate or transparent in their decisions and if they have enough time and skill to use this more cognitively demanding rule.

The distinction between normative and descriptive approaches is somewhat overdrawn because current normative models such as SEU are at least partially descriptive. That is, the SEU model is defined in terms of subjective parameters (i.e., subjective probability and utility) that can be said to be descriptive because they are free parameters that are specific to that DM. By contrast, the earlier EV model was unambiguously normative in the sense that its parameters were fixed (i.e., not estimated from the DM) once the nature of the event had been

determined. The contrast between normative and descriptive models is further muddled by the fact that some models of decision-making labeled descriptive have an *implicit* normative quality. The [Janis and Mann \(1977\)](#) conflict-theory model, for example, contains a flow chart that describes a sequence of steps through which a DM responds to impending danger. The model's normative character can be discerned from the authors' use of terms such as "unbiased assimilation of new information," "effective planning," "defective decision-making," and "maladaptive behavior."

To say that the distinction between normative and descriptive models is fuzzy is quite different from saying that it is altogether useless. Instead, current JDM models are more appropriately interpreted as differing in their locations on a continuum, rather than by membership in distinctive categories. At the most purely normative end of the continuum are mathematical programming algorithms whose inputs are actuarially determined probabilities, objectively defined goal functions, and clearly described constraints. At the most descriptive end of the continuum are flow models, such as [Janis and Mann \(1977\)](#), [Fishbein and Ajzen \(1975, 2010\)](#), and [Lindell and Perry \(1992, 2004, 2012\)](#), as well as information processing models such as [Payne, Bettman, and Johnson \(1993\)](#).

### III RESEARCH ON CRITICAL DECISION ELEMENTS

Studies since the 1950s have generated a large volume of findings on all aspects of JDM research. This section briefly summarizes some major research findings and the experimental methods used to assess four critical decision elements—generation of alternatives, subjective probability judgments, utility judgments, and search for additional information.

#### A Generation of Alternatives

The EV model and its successor, the SEU model, are narrow models of *choice* rather than broad models of *decision-making* because they presume that a DM has a readily accessible set of decision alternatives with well-defined attributes. Accordingly, many studies explicitly present DMs with a set of predetermined alternatives having predefined attributes, as in the prototypical consumer decision involving the choice among multiple cameras on a store's shelves ([Edwards & Fasolo, 2001](#)). For example, [Payne's \(1976\)](#) classic study of information search examined DMs' requests for information about a set of one-bedroom apartments characterized by quantitative (e.g., rent) and qualitative (e.g., noise level) attributes. However, the way in which alternatives are generated—described by others, retrieved from memory, or constructed *de novo*—can significantly affect decisions. In a laboratory experiment, [Hertwig, Barron, Weber, and Erev \(2004\)](#) found that decisions based on descriptions of alternatives relied more on rare events than did decisions based on information retrieved from memory, presumably because the descriptions made information about these rare events easier

to maintain in working memory as information about the options was processed cognitively. [Hertwig and Erev \(2009\)](#) attributed this description–experience gap to inadequate sample sizes, overweighting of recent information, underestimation of the frequency of rare events, and differences in the heuristics by which the two types of information were processed.

Curiously, there has been little laboratory research on the generation of novel decision alternatives, which is important because [Anderson \(1983\)](#) found that DMs in the Cuban missile crisis generated alternative courses of action by an iterative process involving the generation of an initial option, identifying a flaw in that option (i.e., an important attribute on which it performed unacceptably), generating a new option that avoided the flaw, re-evaluating the available options in terms of the new set of attributes, and continuing the process until an acceptable course of action was identified. [Engleman and Gettys \(1985\)](#) found that domain-general strategies and specific knowledge both contribute to high performance in generating decision alternatives in familiar problems because they provide access to existing schemas. Unfamiliar problems are more cognitively demanding because they require schema construction ([Gettys, Pliske, Manning, & Casey, 1987](#)). Although this line of research does not appear to have been pursued by JDM researchers, relevant research can be found in theories of creative problem solving such as [Hélie and Sun's \(2010\) explicit–implicit interaction \(EII\) theory](#), which proposes that DMs must draw upon both explicit and implicit knowledge when solving novel problems. Hélie and Sun argue that EII theory can account for important findings from research on two critical problem-solving stages: incubation (a period of conscious and unconscious rumination) and insight (the typically rapid achievement of a solution).

## B Subjective Probability Judgments

One way DMs can construct subjective probability judgments is by retrieving instances from their memories of personal experiences. Although the *frequency* that relevant events have occurred influences these judgments, they are also influenced by these events' *recency*, *salience*, and *memorability* ([Kynn, 2008](#)). In addition, DMs can construct subjective probabilities by means of implications from other beliefs. For example, a DM who believes "new technology is likely to be dangerous" and "nanotechnology is a new technology" will probably infer that "nanotechnology is likely to be dangerous."

Another source of subjective probability judgments is verbal, numeric, or graphic communication from others (e.g., "There is a 60% chance of snow in the next 24 hours"). Many studies have examined the communication of uncertainties about the risk of exposure to a wide range of environmental pollutants, diseases, and medical treatments ([Kurz-Milcke, Gigerenzer, & Martignon, 2008](#); [Visschers, Meertens, Passchier, & de Vries, 2009](#)). Moreover, a group of experimental studies have investigated the effects of providing graphical representations of uncertainties ([Ancker, Senathirajah, Kukafka, & Starren, 2006](#);

(Bisantz, Marsiglio, & Munch, 2005). These studies have produced a significant body of evidence that, for example, DMs tend to confuse relative risk reduction for absolute reduction, relative frequency graphs are more likely to be interpreted correctly than numerical probabilities, DMs are confused by changes in denominators (e.g., is 1 chance in 7 greater than 4 chances in 33?), results of screening tests can be confusing because false-positives and their effects are not readily understood, and there are widely differing interpretations of verbal probability labels.

A significant problem with subjective probability judgments is that DMs sometimes produce numbers that do not conform to the requirements of probability theory. For example, Fox and Tversky (1988) reported that although the median sum of probabilities for two mutually exclusive and exhaustive events was approximately 1.0, the median sum of probabilities for four events was 1.4, and the median sum of probabilities for eight events was 2.4. Others have reported similarly excessive sums of probabilities (Tversky & Koehler, 1994; Wright & Whalley, 1983; Wu, Lindell, Prater, & Samuelson, 2013).

Another deficiency in probability judgments is unrealistic optimism, which Shepperd, Klein, Waters, and Weinstein (2013, p. 396) defined as “a favorable difference between the risk estimate a person makes for him- or herself and the risk estimate suggested by a relevant, objective standard.” Unrealistic absolute optimism involves comparison to an absolute standard such as epidemiological evidence of disease risk, whereas unrealistic comparative optimism involves comparison to other people. The latter might be due to unjustified positive assessments of one’s own risk, unjustified negative assessments of others’ risks, or both. Both absolute and comparative judgments can be analyzed at either the individual or the group level of analysis. DMs are less likely to make unrealistically optimistic judgments when they expect to be held accountable for their judgments (perhaps because they think more carefully about reasons for their judgments), when they have experience with the event (especially when it provides proximal feedback), when they compare themselves with similar rather than dissimilar others (e.g., comparison to “average person” of a specific category such as “men over 65” rather than a nonspecific category such as “the average person”), or when they expect to be able to control the outcome (Shepperd et al., 2013).

DMs are also subject to hindsight bias, which is a tendency to view events as more foreseeable *after the fact* than they actually would have appeared at the time action was taken (Blank, Musch, & Pohl, 2007; Louie, Rajan, & Sibley, 2007). For example, Fischhoff’s (1975, Experiment 1) initial study of this topic provided descriptions of four different events, each of which was followed by four possible outcomes. For all four events, experiment participants were assigned either to a condition in which they were told that one of four outcomes actually occurred (the after conditions) or to a condition in which they were not told about the actual outcome of the event (the before condition). Participants in the after conditions made higher probability judgments of the outcome that

they were told occurred than did the participants in the before condition. Thus, although most DMs recognize it would be unfair to use later information to second guess earlier decisions, research on hindsight bias has shown that cognitive biases can limit DMs' ability to ignore subsequently received information.

Still other deficiencies in subjective probability judgments include insensitivity to base rates (subjective probability judgments tend to be unduly influenced by the similarity of an instance to a relevant category despite the infrequency of that category) and insensitivity to sample sizes (DMs tend to act as if probabilities based on large samples are no more stable than probabilities based on small samples). In addition, DMs tend to be overconfident in the accuracy of their judgments (for more extensive descriptions of judgment deficiencies, see [Bazerman & Moore \(2008\)](#) and [Baron \(2008\)](#)).

## C Utility Judgments

Another significant body of JDM research has revealed framing effects—cases in which a situation described in different ways produces different responses ([Kühberger, 1998](#)). [Tversky and Kahnemann's \(1981\)](#) Asian disease problem provides an example of how decisions in tasks framed in terms of “lives lost” can differ significantly from those in tasks framed in terms of “lives saved.” Specifically, they tested DMs’ choice between two programs to combat an Asian disease that was expected to kill 600 people if no action was taken. Program A would save 200 people for certain, whereas Program B had a 1/3 probability of saving all 600 but a 2/3 chance of saving none. After reporting their choice of Program A versus Program B, DMs chose between two other programs—C, in which 400 people would die for certain, and D, in which there was a 1/3 probability of saving all 600 but a 2/3 chance of saving none. Most (72%) experiment participants preferred Program A to B, but 78% preferred D to C—even though Program C is formally equivalent to (i.e., has the same payoff structure as) Program A and Program D is formally equivalent to Program B. That is, anyone who preferred A to B also “should have” (according to the expected value principle) preferred C to D.

[Wagenaar, Keren, and Lichtenstein \(1988\)](#) demonstrated the robustness of decision framing effects by presenting a single deep structure in terms of 11 different surface structures that varied in presentation, confounding variables, and context. The preference for the certain outcome ranged from 8 to 83% over these 11 surface structures, confirming that few, if any, DMs were able to extract the deep structure of the problem from its surface structure, and that the framing of a decision significantly affected DMs’ choices.

DMs’ difficulty in making judgments on attributes for which they have no absolute standard of evaluation can also be seen in [Hsee’s \(1996\)](#) demonstration that willingness to pay for Dictionary A (which contained 10,000 words and had a pristine cover) was lower than for Dictionary B (which contained 20,000 words and had a slightly tattered cover) only for a group of subjects that was

able to make an explicit comparison between the two. By contrast, a group that rated Dictionary B in isolation was willing to pay more for it than another group that rated Dictionary A in isolation. These results indicate that seeing information about the number of words in Dictionary A provided a reference point for judging the number of words in Dictionary B that offset the minor difference in the quality of the covers.

In cases in which there is a reference point, Kahneman and Tversky's (1979) prospect theory proposes that DMs frame outcomes in terms of gains or losses from that reference point, with a roughly S-shaped evaluation function passing through the origin. In addition, it predicts that the pleasure associated with winning a given amount of money is smaller than the displeasure associated with losing that same amount. Moreover, prospect theory replaces probabilities with decision weights that are larger than the corresponding probabilities for small values (near 0) and smaller than the corresponding probabilities for large values (near 1). The nonlinearities in the evaluation and weighting functions can create different decision frames that produce different choices and, in some cases, preference reversals.

In addition to outcome uncertainty, outcome delay is another factor affecting DMs' judgments. Researchers in developmental psychology have long been interested in children's delay of gratification—a willingness to forego immediate satisfaction in order to achieve temporally remote goals (for a review, see Mischel, Shoda, & Rodriguez, 1988). When outcomes are delayed, many DMs experience adverse effects such as procrastination (Steel, 2007), myopic time horizons that focus only on short-term consequences (Kunreuther, Pauly, & McMorrow, 2013), or display hyperbolic discounting that weights outcomes "at a rate that is a decreasing function of delay, so that each successive unit of delay has a smaller proportional impact than the preceding one" (Read, Frederick, & Airolidi, 2012, p. 177).

## D Search for Additional Information

An important implicit assumption of SEU theory is that search for additional information is motivated solely by concerns about accuracy. DMs are likely to engage in unbiased search (i.e., be guided by accuracy motivation) when their beliefs, attitudes, or behaviors are unrelated to their values; when there is only low-quality information available; when they have low confidence in their positions; or when they are generally open-minded. However, DMs sometimes seek additional information that is attitude-consistent (i.e., be guided by defense motivation), especially when they believe that important beliefs, attitudes, or behaviors to which they are committed are being challenged (Hart et al., 2009). There is evidence that attitude-consistent information is perceived to be higher in quality than attitude-inconsistent information (Chaiken, Giner-Sorolla, & Chen, 1996), so it is perceived information quality that determines preferences for attitude-consistent information (Fischer, Schulz-Hardt, & Frey, 2008).

Fischer and Greitemeyer (2010) summarized an extensive line of research as indicating that accuracy motivation decreases selective exposure only when the accuracy cue is related to the decision-making context (i.e., the quality of the decision). By contrast, accuracy motivation increases selective exposure when the accuracy cue is related to the information search context. This occurs when task conditions lead DMs to bolster the justification for their decisions by selecting information that *appears* to be more valid because, as Chaiken et al. (1996) concluded, it is consistent with their preexisting knowledge.

Other studies have examined selective exposure from the perspective of information avoidance rather than choice between consonant/congenial and dissonant/uncongenial information (Sweeny, Melnyk, Miller, & Shepperd, 2010). For example, Howell and Sheppard (2012) framed the search for information about personal health as a decision task. They found that having DMs think about their reasons for seeking or avoiding information about a medical condition decreased their avoidance of information seeking. Moreover, there was an even lower level of information avoidance among those in the contemplation conditions when the medical condition was described as treatable—that is, when the information actually had some value in determining future actions. The latter finding is consistent with findings by Goodall and Reed (2013), who manipulated the level of certainty about a threat and the efficacy of a protective action. They found that uncertainty about a threat was associated with intentions to *seek* further information, whereas uncertainty about the efficacy of the protective action was associated with intentions to *avoid* further information. Thus, both studies found support for Hart and colleagues' (2009) proposition that accuracy motivation is enhanced when there is information that is perceived to be useful in making decisions that have important personal outcomes.

## IV EFFECTS OF JUDGMENT HEURISTICS

Some JDM researchers have sought to account for judgment deficiencies by attributing them to *heuristics*, which Tversky and Kahneman (1974, p. 1124) defined as “principles which reduce the complex tasks of assessing probabilities and predicting values to simpler judgment operations [that] are quite useful, but sometimes they lead to severe and systematic errors.” In a summary of their own and others’ JDM experiments, Tversky and Kahneman (1974) identified three heuristics—representativeness, availability, and adjustment and anchoring. The *representativeness heuristic* is used to produce judgments of probability by assessing the similarity of an object to the class of which it is a member. Kahneman and Tversky (1973) demonstrated this heuristic by asking experiment participants to judge the proportion of first-year graduate students in each of nine disciplines and then presenting them with a brief description of Tom W., a stereotypical “nerd.” One group of participants was asked to judge the probability that Tom was a student in each of the disciplines, whereas another group was asked to judge how similar Tom was to the

typical student in each discipline. The first group judged Tom to be most likely to be in computer science—even though this group judged the proportion of students in computer science (i.e., the base rate) to be the second lowest of all nine disciplines. The second group judged him to be most like computer science students, suggesting that the first group made its judgments of probability on the basis of similarity to the typical computer science student. There is a logical basis for the representativeness heuristic because classes are defined by similarity among their members. However, this heuristic can lead to errors such as the neglect of base rates (an object is more likely to be from a class with many members than a class with extremely few members, even if it seems more “similar” to the prototypical member of the less frequent class); see [Koehler \(1996\)](#) for a review and [Barbey and Sloman \(2007\)](#) and accompanying commentary for further discussion. Representativeness can also cause insensitivity to sample size (failing to recognize that a deviation from the population likelihood is more likely in a small sample than in a large sample) and other biases.

The *availability heuristic* is used to produce judgments of probability by the ease and frequency with which relevant objects are retrieved from memory, and the *anchoring and adjustment heuristic* is used to produce revised judgments after producing or viewing an initial value. Disturbingly, numbers that are readily accessible in memory—such as the last three digits of DMs’ phone numbers ([Bazerman & Moore, 2008](#))—can influence judgments even when DMs know these numbers are completely irrelevant.

Research on judgment heuristics and their consequent biases initially received a skeptical reception from economists committed to the assumption that DMs make decisions according to the EV model. Indeed, [Grether and Plott \(1979\)](#) assumed that violations of the EV model were created through artificial laboratory conditions, so they sought to challenge a study of preference reversals—situations in which DMs more frequently chose gambles with a higher probability to win but considered gambles with higher monetary payoffs to be worth more ([Lichtenstein & Slovic, 1971, 1973](#)). To Grether and Plott’s surprise, their laboratory studies were unsuccessful in explaining away Lichtenstein and Slovic’s results—a powerful demonstration of the usefulness of laboratory experiments to test theoretical assumptions.

[Hogarth \(1981\)](#) acknowledged the biases that [Tversky and Kahneman \(1974\)](#) and other researchers had reported but called attention to two issues needed to interpret those findings. First, he contended that most of the studies involved discrete laboratory tasks even though many situations involve continuous judgments that allow DMs to minimize errors by revising preliminary judgments using task feedback (although the anchoring and adjustment heuristic and the hindsight bias raise questions about DMs’ ability to consistently utilize feedback effectively). Second, he noted that many tasks outside the laboratory provide ambiguous or delayed feedback, so it is important to examine the extent to which these tasks also have characteristics that frustrate accurate judgments.

Since then, other researchers have examined the generalizability of Kahneman and Tversky's heuristics, as well as those discovered by other researchers (for summaries of heuristics and biases, see [Bazerman & Moore \(2008\)](#) and [Gilboa \(2011\)](#)). Some evidence of heuristic use outside the laboratory can be seen in a variety of domains, such as retirement savings ([Benartzi & Thaler, 2007](#)), finance ([Muradoglu & Harvey, 2012](#)), accounting ([Smith & Kida, 1991](#)), judicial decisions ([Rachlinski, 2000](#)), business ventures ([Busenitz & Barney, 1997](#)), information technology ([Browne & Parsons, 2012](#)), medicine ([Elstein, 1999](#)), and even romantic relationships ([Joel, MacDonald, & Plaks, 2013](#)), suggesting that these heuristics are not simply laboratory artifacts. However, some of these authors have expressed reservations about the pervasiveness and impact of judgment heuristics. For example, [Schwartz and Elstein \(2009\)](#) remarked:

*Behavioral decision research conducted to date has been concerned primarily with demonstrating that a particular phenomenon exists.... More research is needed to assess the prevalence of these errors and to determine how often treatment choices are affected by diagnostic errors caused by these biases.* (p. 251)

On the other hand, some researchers have found professionals who were even more susceptible to judgment heuristics than the students used in many JDM samples (e.g., for a summary of studies in behavioral finance, see [Muradoglu & Harvey, 2012](#)).

Gigerenzer and colleagues challenged heuristics and biases research by articulating a *fast and frugal heuristics* (FFH) approach ([Gigerenzer et al., 1999](#); [Gigerenzer & Gaissmaier, 2011](#)), which proposes that DMs choose a specific heuristic to achieve a task goal in a given context. Unlike the heuristics and biases paradigm, the FFH approach defines a heuristic as “a strategy that ignores part of the information, with the goal of making decisions more quickly, frugally, and/or accurately than more complex methods” ([Gigerenzer & Gaissmaier, 2011, p. 454](#)). FFH rejects the assumption that there is a strong trade-off between speed and accuracy—finding that fast heuristics can sometimes achieve an outcome that is almost as accurate as a substantially more time-consuming normative process such as multiple regression analysis.

[Reyna \(2004, p. 61\)](#) described an alternative critique of both the heuristics and biases and FFH approaches by noting that “[r]easoning is independent of memory ... [and] ... errors in reasoning occur on tasks that impose few demands on memory capacity.” Instead, she summarized an extensive series of studies on *fuzzy-trace theory* (FTT) that suggest DMs begin to process judgment problems by generating two ways of encoding—verbatim (specific words or numbers) and gist (the general meaning of a piece of information). In addition, they recall relevant heuristics from memory and apply those heuristics to their encoded representation of a problem. Judgment biases may be due to errors in any one of those three stages of processing. In one test of the effects of verbatim and gist representations, [Reyna and Brainerd \(1991, Experiment 1\)](#) presented

experiment participants with three variants of Tversky and Kahneman's (1981) Asian disease problem. In an *outcomes deleted* version, they substituted the vague term "some" for the specific numerical (i.e., verbatim) outcomes "200 saved" and "400 die" and the vague term "many" for the specific numerical outcome "600 saved." In a *probabilities deleted* version, they substituted the vague terms "some" and "higher" for the specific probabilities "1/3" and "2/3," respectively. In a third version, they provided only vague versions of both outcomes and probabilities. The results confirmed their prediction that preference reversals due to framing effects could be explained by DMs' encoding of qualitative gist rather than verbatim numerical encoding.

## V EFFECTS OF TASK AND COGNITIVE CONSTRAINTS

Another line of JDM research has sought to identify a broader set of sources for judgment deficiencies. Similar to Peterson and Beach's (1967) rather positive conclusion about JDM performance, this line of research has noted that judgment deficiencies are pervasive but do not seem to be universal. For example, weather forecasters seem to provide well-calibrated subjective probabilities; their judgments of the probability of precipitation are generally quite consistent with the percentage of days on which there is measurable precipitation. Indeed, some researchers have attributed many JDM deficiencies to task and cognitive constraints.

### A Task Constraints

Judgments and decisions are made in situations that vary in their uncertainty, complexity, and time constraints. *Uncertainty* arises from imperfect (i.e., incomplete or unreliable) information about variables on which outcomes depend. Thus, DMs often make a "best guess" but also recognize that this best guess will not be perfectly accurate (Hastie, 2001).

Decision *complexity* arises from the number of alternative actions; the number, metric quality, commensurability, and intercorrelations (redundancy) of the attributes; the number, metric quality, and commensurability of outcome variables; and the compatibility of attributes and responses (judgments or decisions). Moreover, Karelaiia and Hogarth (2008) concluded that task characteristics impairing judgment include a large number of cues; intercorrelations among cues, especially when the cues have negative corrections with each other and their correlations with the criterion have differential weights and signs; "achieved cues" that are inferred by the DMs rather than given to them; complex functional relations between cues and criterion (e.g., U-shaped and inverted U-shaped); and low levels of predictability of the criterion variable from the cues. In addition, judgments and decisions are more complicated in dynamic situations than in static situations, such as when DMs are deciding how to respond to an approaching hurricane that might or might not strike their

jurisdictions (Wu, Lindell, & Prater, 2013) or are trying to control a system that has complex multistage causal relationships involving feedback relations and time lags (Brehmer, 1999; Rigas & Brehmer, 1999).

*Time constraints* take two forms—urgency and delay. *Urgency* arises from the immediacy and importance of outcomes that will result if appropriate action is not taken (Hammond, 2000). Urgency can explicitly constrain the search for alternative courses of action by reducing the amount of time available for information search and also impair a DM's ability to process the information that is available. As noted previously, *delay* occurs when a decision's outcomes will not be realized until long after the decision is made, as is the case with retirement investments.

## B Cognitive Constraints

Research on judgment heuristics has attributed the use of these shortcuts to cognitive constraints, but little of this research has systematically examined how JDM processes are affected by limits on attention, working memory, and the ability to retrieve information from long-term memory. The initial impetus for addressing this issue in JDM research was Simon's (1957) concept of *bounded rationality* as a reason why DMs *satisfice*—that is, settle for the first acceptable solution. Indeed, just as decision performance has been compared to that of the rational economic DM, so too has information processing performance been compared to that of a computer (Hunt, 1971). The concept of a *cognitive miser* (Taylor, 1980) arises from bounded rationality; given their cognitive constraints, as well as the multiple decision problems they might face in the course of their daily lives, DMs are often motivated to minimize the amount of effort they devote to any given decision. However, DMs do not invariably minimize decision effort; JDM researchers have proposed models of contingent decision-making in which DMs vary their decision strategies in response to the relative importance of accuracy versus time and effort (Beach & Mitchell, 1978; Payne et al., 1993), although Gigerenzer and associates (Gigerenzer et al., 1999; Gigerenzer & Gaissmaier, 2011) contend that some decision strategies can be both fast and accurate.

One basic finding from laboratory research is that JDM processes take place in working memory where they are influenced both by information retrieved from long-term memory and by incoming information that is heeded in the external environment (Fiske & Taylor, 1991; Matlin, 2009). Long-term memory influences what information is heeded or sought in the external environment (“top-down” processing), and incoming information influences what is retrieved from long-term memory (“bottom-up” processing).

### 1 Long-Term Memory

Long-term memory is often thought to be a warehouse into which information is transported and from which it is subsequently retrieved, but it is increasingly

recognized that memory is a representational process in which incoming information is encoded in terms of its compatibility with existing schemas as well as being recorded verbatim (Koriat, Goldsmith, & Pansky, 2000). As a result of constructive and reconstructive processes, DMs can fail to distinguish actual from imagined events, misattribute statements to the wrong speaker, report false memories, or fail to recall details of events that occurred. For example, the wording of questions asked at the time of retrieval can change DMs' judgments, as Loftus and Palmer (1974) demonstrated in a laboratory experiment that found that DMs who were asked how fast two cars were going when they *smashed* each other estimated greater vehicle speeds than those who were asked how fast the cars were going when they *contacted* each other. However, the accuracy of recall is enhanced if DMs can accurately judge the probability of each item being correct and can either screen out items in which they have low confidence or increase the level of abstraction at which they report. For example, a DM's confidence that a pedestrian was struck by a Porsche is likely to be lower than that he was struck by a sports car and is likely to be lower than he was struck by a sports car than by an automobile (rather than a truck or bus). Retrieval is also more accurate when the conditions at that time are more similar to the conditions when the material was first learned (Proctor & Vu, 2012).

Many researchers distinguish episodic memories about personally experienced events from procedural memories about how to perform tasks and semantic memory, which is organized knowledge about the world (e.g., Schacter, 1999; Tulving, 2002). Task expertise is a crucial aspect of semantic memory because tasks about which DMs have learned by extensive experience or received information from others provide the basis for schemas or mental models about specific categories of objects and events that are stored in long-term memory. *Schemas* (which generally refer to representations of static systems; Taylor & Crocker, 1981) and *mental models* (which generally refer to the representations of dynamic systems; Moray, 1999) are generic knowledge structures that organize the information a DM uses to understand and generate inferences about a specific knowledge domain (Bostrom, Fischhoff, & Morgan, 1992). Schemas (a term that is often used broadly to include mental models, stereotypes, scripts, implicit theories, and lay theories; see Fiske & Taylor, 1991) comprise concepts, their defining attributes, and their relationships that guide attention to incoming information, classify situations, guide the search for additional information, and supply missing information from long-term memory that is not available externally. Because a schema is a DM's mental representation of a decision problem, it can contain deficiencies and erroneous beliefs as well as accurate information.

## 2 Attention

Attention plays a crucial role in allocating mental resources, which are typically considered to be limited to approximately four to six independent variables (Strayer & Drews, 2007a). DMs' preoccupation with some aspects of a situation

can cause *inattentional blindness*, in which they fail to notice critical cues that are outside the focus of attention (Strayer & Drews, 2007a). For example, laboratory simulations have shown that DMs are generally unable to effectively process information about one task (e.g., driving) while performing another, such as talking on a cell phone (Strayer & Drews, 2007b). However, the situation is complicated by the fact that attention relies on multiple resources (Wickens & McCarley, 2008) involving processing of information at different stages (perception, cognition, and response) using different codes (verbal and spatial). In turn, the perceptual stage has two different modalities (visual and auditory), and the visual modality has two different channels (focal and ambient vision). According to this model, concurrent processing of two tasks will be more effective to the extent that those tasks utilize resources from different levels of these dimensions (Wickens, 2008). When task demands exceed mental resources (i.e., they create excessive mental workload; Pickup et al., 2005), a DM can (but not necessarily *will*) choose to reallocate resources from one task to another—because of either external interruptions or explicit decisions about the relative importance of the tasks. An important implication of the multiple resources model is that DMs can process, perhaps without explicit awareness, information that is peripheral to the information on which they are focusing. More generally, however, DMs must selectively and flexibly allocate attention in complex situations when the number of variables exceeds their mental workload. Moreover, information displays should be designed to accommodate DMs' attention limits.

### 3 Working Memory

Working memory is thought to include a central executive that performs computational processes as well as coordinating components that process verbal, visual, and spatial information (Baddeley, 2000, 2012). It provides DMs with an ability to actively maintain task-relevant information while performing a cognitive task (Boduroglu, Minear, & Shah, 2007); however, it is limited in capacity, so DMs try to overcome this limit by encoding and organizing (“chunking”) incoming information on the basis of existing schemas (Ericsson & Kintsch, 1995). Thus, researchers have shown that chess masters have a better memory than novices for meaningful chess positions but are no different from novices in their memory for random distributions of chess pieces (Chase & Simon, 1973). As is the case with attention, the design of effective information displays can cope with the limitations of DMs’ working memory by reducing the amount of information, reducing the difficulty of integrating that information, or reducing the need to mentally transform that information by providing it in a form that is consistent with a DM’s schemas (Boduroglu et al., 2007).

## VI DEEPER FOUNDATIONS OF JDM PROCESSES

In recent years, there has been increased attention to two new perspectives on JDM research. Expanding on distinctions made in theories of memory, some

researchers have proposed that judgments and decisions are the products of two (or possibly more) distinct cognitive systems. A closely related argument is that emotion, which has long been recognized as an essential element of attribute utilities and outcome valences, affects judgments and decisions in other—more direct—ways as well.

## A Dual-System Models of Judgment

A number of dual-system (also called dual-process; [Evans, 2008](#)) models have proposed the existence of two distinctly different ways of processing internal information to produce judgments. For example, [Kahneman's \(2003\)](#) System 1 is defined by intuitions about beliefs and preferences that are readily accessible (although their sources are inaccessible) because of repeated practice. By contrast, System 2 is defined by conscious thought that is slower, more effortful, and more open to introspection, reflection, explanation, and alteration. [Smith and DeCoster \(2000\)](#) proposed two memory systems—a slow-learning associative system based on the accumulation of instances and a fast-learning system based on the acquisition of rules. Although these are by no means the only dual-process models (for a list, see [Evans, 2008](#)), they provide an indication of the range of assumptions involved. According to [Kruglanski and Orehek \(2007\)](#) and [Evans \(2008\)](#), dual-system models generally agree that one system—what [Epstein \(1994\)](#) calls the experiential system—involves an instance-based associative process that automatically produces rapid, effortless inferences about whose sources a DM has no insight or ability to explain (i.e., a “gut feeling”). By contrast, the other system—what [Hammond \(1996\)](#) calls the analytic system—comprises abstract propositions that DMs can receive directly from others as well as learn by reflection on their own experience, can consciously and effortfully retrieve and manipulate, and can explain to others.

The analytic system, with its connection to abstract propositions, might seem to be a defining characteristic of expertise and superior judgment performance, but that is not necessarily true. [Klein's \(2008\)](#) research on some types of experts reveals that they do not consciously retrieve and manipulate abstract propositions but instead recognize a familiar situation, automatically retrieve the most suitable action, evaluate its likely effectiveness, and then implement it. The more complex analytic processing associated with the classic decision model is undertaken only in unfamiliar situations. Indeed, [Klein's \(2008\)](#) recognition-primed decisions are somewhat similar to habits in their influence on behavior. According to [Neal, Wood, and Quinn \(2006, p. 198\)](#), habits are “automated response dispositions that are cued by aspects of the performance context (i.e., environment, preceding actions).” Habits form in response to attempts to achieve goals and are strengthened by repetition; over time, however, the original goals may lose their impact on the continued behavior ([Neal, Wood, Wu, & Kurlander, 2011](#)). Habits are pervasive; diary studies show DMs repeat

approximately 45% of their behaviors in the same locations each day (Wood, Quinn, & Kashy, 2002). Moreover, DMs make conscious decisions about their activities only when habit strength is low or when changed contexts remove the cues to habitual behavior. Thus, repetitive behaviors are more resistant to change by persuasion and other interventions directed toward changing DMs' beliefs about decision elements.

For example, Neal et al. (2011, Experiment 1) conducted a field experiment in which participants were randomly assigned to a two (locations—a theater to watch movie trailers or a meeting room to watch music videos) by two (popcorn freshness—1-hour or 7-day-old) design. Upon arriving at the designated location, participants completed a questionnaire about their current level of hunger and the length of time since their most recent meal. After completing the questionnaire, they were given a box of popcorn and a bottle of water and viewed 15 minutes of trailers for six unreleased films. When the trailers finished, the participants rated their liking of their popcorn and reported their frequency of eating popcorn in theaters, and the experimenters collected the popcorn boxes to determine what percentage of the popcorn had been consumed. Participants in the theater (which was assumed to be an environmental cue for eating popcorn) ate the same amount of stale as fresh popcorn if they had strong popcorn-eating habits and significantly less popcorn if they had weak habits. Those in the meeting room (a setting that lacked environmental cues for eating popcorn) ate less stale popcorn than fresh regardless of popcorn-eating habits. Experiment 2 was also conducted in a theater but manipulated the hand with which the participant ate popcorn (dominant or nondominant). The results of two experiments suggested that disruption of habit automaticity by being in an uncued location or eating in an awkward way brought participants' eating behavior under their intentional control and, thus, they were more responsive to the quality of what was being eaten.

## B Emotions and Judgment

The classical way of addressing the role of emotion in normative JDM models is through the utilities associated with each of the attributes. These are *anticipated* emotions that a DM expects to result from the consequences of a decision (Loewenstein, Weber, Hsee, & Welch, 2001). However, judgments and decisions can also be affected by emotions that are concurrent with the decision process. These *anticipatory* emotions might be prompted by thoughts that are directly associated with the decision process, by characteristics of the decision context such as urgency (e.g., Hammond, 2000), or by cues that are completely irrelevant to the decision (Slovic, Finucane, Peters, & MacGregor, 2002). In some cases, the influence of affect on judgments completely bypasses explicit awareness (Loewenstein et al., 2001). Slovic et al. (2002) contended that affect influences decisions because it is linked to images of objects and events represented in long-term memory. They noted that extraneous affect such as

music in movies and smiling models in advertisements are calculated attempts to increase the amount of positive affect and, thus, change decisions.

Other researchers have contended that some decisions are made substantially on the basis of emotion that is associated with personal experience. For example, Epstein (1994, p. 716) asserted that “[t]he experiential system is assumed to be intimately associated with the experience of affect.” Zajonc (1980) made an even stronger statement by claiming that DMs’ first reactions to objects are automatic affective reactions and that DMs’ claims of rational decisions are post hoc rationalizations. Specifically, “[w]e do not just see ‘a house’: we see a handsome house, an ugly house, or a pretentious house” (Zajonc, 1980, p. 154). However, the validity of this assertion depends on the object being judged, the DM’s experience with that type of object (as encoded in his or her schema for that class of objects), and the judgment being made about the object. A judgment of a house’s beauty might be immediate for most DMs, but this is not the case for a judgment of its affordability, which must be computed rather than directly perceived.

To summarize the research on affect in JDM, Loewenstein et al. (2001) proposed a four-stage model in which anticipated outcomes (including anticipated emotions), subjective probabilities, and other factors such as vividness and background mood (the first stage) affect both cognitive evaluations and feelings (the second stage), which in turn affect behavior (the third stage) and, ultimately, actual outcomes including emotions (the fourth stage). Thus, “feelings can also arise without cognitive mediation” and “the impact of cognitive evaluations on behavior is mediated, at least in part, by affective responses” (Loewenstein et al., 2001, p. 271). They also cited Slovic et al. (1991) as indicating that ratings of the valences of city and state attributes generated by a free-response method provide measures of pure affect. However, the claim that attributes generated in this procedure are devoid of any cognitive content conflicts with Fishbein and Ajzen’s (1975) theory of reasoned action, which refers to these measures as *salient beliefs* that, as the name implies, are measures of cognition as well as affect.

## VII SUMMARY AND CONCLUSIONS

Decades of laboratory and field experiments have provided evidence of systematic deviations from normative JDM principles. In some respects, it is unremarkable that DMs seem to fare poorly with respect to normative models because even the oldest of them—the EV principle—is less than 300 years old; Bayes’ theorem is only approximately 250 years old, and the correlation coefficient is only approximately 125 years old. The fact that these normative JDM models were proposed long after such analytic principles as the Pythagorean theorem, which is at least 2,500 years old, should be evidence of their non-intuitive nature. A major source of judgment biases is the fact that judgment is often comparative rather than absolute, especially for people who lack expertise

regarding a specific situation. EU theory requires absolute judgments, but DMs use labile reference points because they cannot recall everything stored in long-term memory and cannot simultaneously process in working memory all that they can recall from long-term memory.

DMs appear to use JDM heuristics for motivational as well as cognitive reasons. In some cases, DMs are justified in using heuristics rather than the normative statistical concepts or procedures because the heuristics are faster and require less effort.

The concept of heuristics as proposed by [Tversky and Kahneman \(1974\)](#) has been extremely influential in JDM research since the mid-1970s because it has provided a unifying explanation for a diverse set of judgment deficiencies. Nonetheless, the heuristics and biases program has been criticized on theoretical grounds by [Einhorn and Hogarth \(1981\)](#), [Gigerenzer \(1991\)](#), [Keren and Teigen \(2004\)](#), and others—although not without response (e.g., [Kahneman & Tversky, 1996](#)). Moreover, some of the evidence for the use of heuristics by subject matter experts in their domains of expertise is anecdotal in nature (e.g., [Mileti, 1999](#)). That is, a number of review articles have reported behavior that is “consistent with” heuristics rather than rigorously attributable to them. In some cases, the use of heuristics has been attributed to an experiential system that is distinct from an analytic system. Based on the research to date, it is unclear if there are indeed multiple mental systems rather than a single mental system or a cognitive continuum; see [Evans and Stanovich \(2013\)](#) and the accompanying commentary. Whether or not there are distinct experiential and analytic systems, it is clear that personal experiences and abstract propositions have different effects on JDM processes. An important challenge will be to conduct experiments on the ways in which each of these sources contributes to JDM processes in different types of tasks and for people with different cognitive capacities, especially expertise.

A related issue concerns the role of emotion on JDM processes; SEU theory implicitly assumes that these effects are completely encompassed within anticipated utilities for potential outcomes. Anticipatory emotion experienced at the time a decision is made is usually assumed away, for example, by decision theorists who exclude utility for gambling itself as a determinant of preferences for uncertain outcomes. The relative influence of anticipated and anticipatory emotions probably varies from one person to another (across situations), one situation to another (across persons), and one situation to another for a given person. Although there is agreement on the general principle that anticipatory emotion influences JDM, the specific mechanisms by which this occurs remain unsettled.

One logical question arising from the research on heuristics and biases is why, if our JDM processes are as bad as the research seems to indicate, are we not extinct as a species? The answer seems to be that there is also ample evidence—as early as [Peterson and Beach \(1967\)](#) and as recent as [Karelaia and Hogarth \(2008\)](#)—that DMs can learn from experience and from each

other. Moreover, DMs' cognitive limitations appear to have some adaptive value because, for example, blocking—a piece of information's temporary inaccessibility—avoids the confusion that could result if all memory traces were equally accessible (Schacter, 1999).

As a practical matter, there seem to be four ways to address the deviations of actual JDM processes from normative models. The first way is to replace the DM, which is the solution many large corporations use for complex scheduling and inventory problems. Replacing the DM can yield very rapid, high-quality decisions but typically requires a major investment of time and money to install and maintain. The second way is to train DMs by explaining the normative models and the situations to which they apply. This, of course, is the purpose of courses in research methods and statistics, but it is also addressed in courses on other subject matter that involves specific types of decisions, such as medical diagnosis (Schwartz & Elstein, 2009) and personnel selection (Schmitt, 2012). The third way is to design situations to reduce the impact of judgment errors. For example, Keller, Cokely, Katsikopoulos, and Wegwart (2010) describe the development of decision aids for physicians who must determine how to treat possible heart attack victims as they arrive at an emergency room. Finally, the fourth way is to develop models that more accurately describe DMs' actual processes, as is the case for JDM researchers who have proposed nonadditive probability weighting functions (see Tversky & Kahneman (1992) for a description of cumulative prospect theory and see Birnbaum (2008) for some alternatives). Better models of actual JDM processes are important because it is not possible to avoid all suboptimal decisions through the first three solutions. Thus, accurate models of JDM processes are needed to forecast when these will produce substantially worse outcomes than aided JDM processes or normative models.

This review reveals some limitations in the JDM research conducted to date. One general problem is that there are varying definitions of related phenomena such as subjective probabilities/expectancies/instrumentalities, attributes/outcomes, utilities/valences, affect, and familiarity. To date, few researchers have provided psychometric evidence of the construct validity (i.e., reliability and convergent and discriminant validity) of these constructs. Moreover, although experimental research in JDM has called attention to biases (i.e., systematic errors), research on personnel selection (Ryan & Ployhart, 2013) has also been concerned about random errors. Random errors do not, as some might argue, "cancel out." As Table 18.1 indicates, false-positive and false-negative errors can have significant costs that add rather than cancel. In some cases, the costs of bias might be small compared to the costs of random error. Consequently, research is needed to assess the relative impacts of bias and random error on JDM processes.

Another general problem is that despite some tentative efforts, JDM research lacks a comprehensive taxonomy of situations, tasks, and person characteristics that can be used to systematically examine the generalizability of different JDM processes. A notable exception is research examining the correlations of

individual difference measures with the use of heuristics (Stanovich, West, & Toplak, 2010). A related issue concerns DMs' ability to recognize the situations in which each normative principle is relevant and to determine which heuristic should be used. Extension of the research on contingent decision-making (Beach & Mitchell, 1978; Payne et al., 1993) and FFH (Gigerenzer et al., 1999; Gigerenzer & Gaissmaier, 2011) is needed.

Finally, the connections of narrow JDM to broad JDM are relatively weak. The weak connections from narrow JDM to specific problem areas such as medical diagnosis and personnel selection are somewhat understandable, but the weak connections to social cognition, attitudes and behavior, and especially fundamental cognitive processes of attention and memory are more difficult to explain. Considerably more work needs to be done in these areas if JDM research is to progress.

## ACKNOWLEDGMENTS

This research was supported by the National Science Foundation under grants CMMI 0838654 and IIS 1212790. None of the conclusions expressed here necessarily reflects views other than those of the author.

## REFERENCES

- Ancker, J. S., Senathirajah, Y., Kukafka, R., & Starren, J. B. (2006). Design features of graphs in health risk communication: A systematic review. *Journal of the American Medical Informatics Association*, 13, 608–618.
- Andersen, S. M., Moskowitz, G. B., Blair, I. V., & Nosek, B. A. (2007). Automatic thought. In E. T. Higgins, & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 138–175) (2nd ed.). New York: Guilford.
- Anderson, P. A. (1983). Decision making by objection and the Cuban missile crisis. *Administrative Science Quarterly*, 28, 201–222.
- Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Science*, 4, 417–423.
- Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, 63, 1–29.
- Barbey, A. K., & Sloman, S. (2007). Base-rate respect: From ecological rationality to dual processes. *Behavioral and Brain Sciences*, 30, 241–254.
- Baron, J. (2008). *Thinking and deciding* (4th ed.). New York: Cambridge University Press.
- Bazerman, M. H., & Moore, D. A. (2008). *Judgment in managerial decision making* (7th ed.). Hoboken, NJ: Wiley.
- Beach, L. R., & Mitchell, T. R. (1978). A contingency model for the selection of decision strategies. *Academy of Management Review*, 3, 439–449.
- Benartzi, S., & Thaler, R. H. (2007). Heuristics and biases in retirement savings behavior. *Journal of Economic Perspectives*, 21, 81–104.
- Birnbaum, M. H. (2008). New paradoxes of risky decision making. *Psychological Review*, 115, 463–501.
- Bisantz, A. M., Marsiglio, S. S., & Munch, J. (2005). Displaying uncertainty: Investigating the effects of display format and specificity. *Human Factors*, 47, 777–796.

- Blank, H., Musch, J., & Pohl, R. F. (2007). Hindsight bias: On being wise after the event. *Social Cognition*, 25, 1–9.
- Boduroglu, A., Minear, M., & Shah, P. (2007). Working memory. In F. T. Durso, R. S. Nickerson, S. T. Dumais, S. Lewandowsky, & T. J. Perfect (Eds.), *Handbook of applied cognition*. (pp. 55–82) (2nd ed.). Hoboken, NJ: Wiley.
- Bostrom, A., Fischhoff, B., & Morgan, M. G. (1992). Characterizing mental models of hazardous processes: A methodology and an application to radon. *Journal of Social Issues*, 48, 85–100.
- Brehmer, B. (1999). Reasonable decision making in complex environments. In P. Juslin, & H. Montgomery (Eds.), *Judgment and decision making: Neo-Brunswikian and process-tracing approaches* (pp. 9–21). Mahwah, NJ: Erlbaum.
- Browne, G. J., & Parsons, J. (2012). More enduring questions in cognitive IS research. *Journal of the Association for Information Systems*, 13(12), Article 2.
- Brunswik, E. (1952). *The conceptual framework of psychology*. Chicago: University of Chicago Press.
- Busenitz, L. W., & Barney, J. B. (1997). Differences between entrepreneurs and managers in large organizations: Biases and heuristics in strategic decision-making. *Journal of Business Venturing*, 12, 9–30.
- Chaiken, S., Giner-Sorolla, R., & Chen, S. (1996). Beyond accuracy: Defense and impression motives in heuristic and systematic information processing. In P. M. Gollwitzer, & J. A. Bargh (Eds.), *The psychology of action: Linking cognition and motivation to behavior* (pp. 553–578). New York: Guilford.
- Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, 4, 55–81.
- Clemen, R. T., & Reilly, T. (2004). *Making hard decisions with decision tools*. Belmont, CA: Brooks/Cole.
- Cronbach, L. J., & Gleser, G. C. (1957). *Psychological tests and personnel decisions*. Urbana, IL: University of Illinois Press.
- Edwards, W. (1961). Behavioral decision theory. *Annual Review of Psychology*, 12, 473–498.
- Edwards, W., & Fasolo, B. (2001). Decision technology. *Annual Review of Psychology*, 52, 581–606.
- Einhorn, H. J., & Hogarth, R. M. (1981). Behavioral decision theory: Processes of judgment and choice. *Annual Review of Psychology*, 32, 53–88.
- Eiser, J., & Stroebe, W. (1972). *Categorization and social judgement*. New York: Academic Press.
- Elstein, A. S. (1999). Heuristics and biases: Selected errors in clinical reasoning. *Academic Medicine*, 74, 791–794.
- Engelmann, P. D., & Gettys, C. F. (1985). Divergent thinking in act generation. *Acta Psychologica*, 60, 39–56.
- Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist*, 49, 709–724.
- Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, 102, 211–245.
- Evans, J. St. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, 59, 255–278.
- Evans, J. St. B. T., & Stanovich, K. E. (2013). Dual process theories of cognition: Advancing the debate. *Perspectives on Psychological Science*, 8, 223–241.
- Fischer, P., & Greitemeyer, T. (2010). A new look at selective-exposure effects: An integrative model. *Current Directions in Psychological Science*, 19, 384–389.
- Fischer, P., Schulz-Hardt, S., & Frey, D. (2008). Selective exposure and information quantity: How different information quantities moderate decision makers' preference for consistent and inconsistent information. *Journal of Personality and Social Psychology*, 94, 231–244.

- Fischhoff, B. (1975). Hindsight ≠ foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance, 1*, 288–299.
- Fishbein, M., & Ajzen, A. (1975). *Beliefs, attitudes, intentions, and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley.
- Fishbein, M., & Ajzen, I. (2010). *Predicting and changing behavior: The reasoned action approach*. New York: Psychology Press.
- Fiske, S. T., & Taylor, S. E. (1991). *Social cognition* (2nd ed.). New York: McGraw-Hill.
- Fox, C. R., & Tversky, A. (1998). A belief-based account of decision under uncertainty. *Management Science, 44*, 879–895.
- Gettys, C. F., Pliske, R. M., Manning, C., & Casey, J. T. (1987). An evaluation of act generation performance. *Organizational Behavior and Human Decision Processes, 39*, 23–51.
- Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond “heuristics and biases.” *European Review of Social Psychology, 2*, 83–115.
- Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology, 62*, 451–482.
- Gigerenzer, G., Todd, P. M. and the ABC Research Group. (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.
- Gilboa, I. (2011). *Making better decisions: Decision theory in practice*. Chichester, UK: Wiley.
- Goodall, C. E., & Reed, P. (2013). Threat and efficacy uncertainty in news coverage about bed bugs as unique predictors of information seeking and avoidance: An extension of the EPPM. *Health Communication, 28*, 63–71.
- Grether, D. M., & Plott, C. R. (1979). Economic theory of choice and the preference reversal phenomenon. *American Economic Review, 69*, 623–638.
- Hammond, K. R. (1996). *Human judgment and social policy*. New York: Oxford University Press.
- Hammond, K. R. (2000). *Judgments under stress*. New York: Oxford University Press.
- Hart, W., Albarracín, D., Eagly, A. H., Brechan, I., Lindberg, M. J., & Merrill, L. (2009). Feeling validated versus being correct: A meta-analysis of selective exposure to information. *Psychological Bulletin, 135*, 555–588.
- Hastie, R. (2001). Problems for judgment and decisionmaking. *Annual Review of Psychology, 52*, 653–683.
- Hastie, R., & Pennington, N. (1995). Cognitive approaches to judgment and decision making. In J. Busemeyer, R. Hastie, & D. L. Medin (Eds.), *Psychology of learning and motivation: Vol. 32*. (pp. 1–31). New York: Academic Press.
- Haynes, C. M. (2002). The effects of context on cascaded-inference evidence evaluation. *Journal of Economic Psychology, 23*, 469–485.
- Hélie, S., & Sun, R. (2010). Incubation, insight, and creative problem solving: A unified theory and a connectionist model. *Psychological Review, 117*, 994–1024.
- Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science, 15*, 534–539.
- Hertwig, R., & Erev, I. (2009). The description-experience gap in risky choice. *Trends in Cognitive Sciences, 13*, 517–523.
- Hogarth, R. M. (1981). Beyond discrete biases: Functional and dysfunctional aspects of judgmental heuristics. *Psychological Bulletin, 90*, 197–217.
- Howell, J. L., & Shepperd, J. A. (2012). Reducing information avoidance through affirmation. *Psychological Science, 23*, 141–145.
- Hsee, C. K. (1996). Elastic justification: How unjustifiable factors influence judgments. *Organizational Behavior and Human Decision Processes, 66*, 122–129.

- Hunt, E. (1971). What kind of computer is man? *Cognitive Psychology*, 2, 57–98.
- Janis, I. L., & Mann, L. (1977). *Decision making: A psychological analysis of conflict, choice, and commitment*. New York: Free Press.
- Joel, S., MacDonald, G., & Plaks, J. E. (2013). Romantic relationships conceptualized as a judgment and decision-making domain. *Current Directions in Psychological Science*, 22, 461–465.
- Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, 58, 697–720.
- Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80, 237–251.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47, 263–291.
- Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions: A reply to Gigerenzer's critique. *Psychological Review*, 103, 582–591.
- Karelaia, N., & Hogarth, R. M. (2008). Determinants of linear judgment: A meta-analysis of lens model studies. *Psychological Bulletin*, 134, 404–426.
- Keller, N., Cokely, E. T., Katsikopoulos, K. V., & Wegwart, O. (2010). Naturalistic heuristics for decision making. *Journal of Cognitive Engineering and Decision Making*, 4, 256–274.
- Keren, G., & Teigen, K. H. (2004). Yet another look at the heuristics and biases approach. In D. J. Koehler, & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making* (pp. 89–109). Malden, MA: Blackwell.
- Klein, G. (2008). Naturalistic decision making. *Human Factors*, 50, 456–460.
- Koehler, J. J. (1996). The base-rate fallacy reconsidered: Descriptive, normative, and methodological challenges. *Behavioral and Brain Sciences*, 19, 1–53.
- Koriat, A., Goldsmith, M., & Pansky, A. (2000). Toward a psychology of memory accuracy. *Annual Review of Psychology*, 51, 481–537.
- Kruglanski, A. W., & Orehek, E. (2007). Partitioning the domain of social inference: Dual mode and systems models and their alternatives. *Annual Review of Psychology*, 58, 291–316.
- Kühberger, A. (1998). The influence of framing on risky decisions: A meta-analysis. *Organizational Behavior and Human Decision Processes*, 75, 23–55.
- Kunreuther, H. C., Pauly, M. V., & McMorrow, S. (2013). *Insurance & behavioral economics: Improving decisions in the most misunderstood industry*. New York: Cambridge University Press.
- Kurz-Milcke, E., Gigerenzer, G., & Martignon, L. (2008). Transparency in risk communication: Graphical and analog tools. *Annals of the New York Academy of Science*, 1128, 18–28.
- Kynn, M. (2008). The “heuristics and biases” in expert elicitation. *Journal of the Royal Statistical Association*, 171, 239–264.
- Lichtenstein, S., & Slovic, P. (1971). Reversal of preferences between bids and choices in gambling decisions. *Journal of Experimental Psychology*, 89, 46–55.
- Lichtenstein, S., & Slovic, P. (1973). Response-induced reversals of preferences in gambling: An extended replication in Las Vegas. *Journal of Experimental Psychology*, 101, 16–20.
- Lindell, M. K., & Perry, R. W. (1992). *Behavioral foundations of community emergency planning*. Washington, DC: Hemisphere.
- Lindell, M. K., & Perry, R. W. (2004). *Communicating environmental risk in multiethnic communities*. Thousand Oaks, CA: Sage.
- Lindell, M. K., & Perry, R. W. (2012). The protective action decision model: Theoretical modifications and additional evidence. *Risk Analysis*, 32, 616–632.
- Lindell, M. K., & Prater, C. S. (2007). A hurricane evacuation management decision support system (EMDSS). *Natural Hazards*, 40, 627–634.

- Loewenstein, G. F., Weber, E. U., Hsee, C. K., & Welch, E. S. (2001). Risk as feelings. *Psychological Bulletin, 127*, 267–286.
- Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior, 13*, 585–589.
- Louie, T. A., Rajan, M. N., & Sibley, R. E. (2007). Tackling the Monday-morning quarterback: Applications of hindsight bias in decision-making settings. *Social Cognition, 25*, 32–47.
- Matlin, M. W. (2009). *Cognition* (7th ed.). Hoboken, NJ: Wiley.
- Mileti, D. S. (1999). *Disasters by design: A reassessment of natural hazards in the United States*. Washington, DC: Joseph Henry Press.
- Mintzberg, H., Raisinghani, D., & Théorêt, A. (1976). The structure of “unstructured” decision processes. *Administrative Science Quarterly, 21*, 246–275.
- Mischel, W., Shoda, Y., & Rodriguez, M. L. (1988). Delay of gratification in children. *Science, 244*, 933–938.
- Moray, N. (1999). Mental models in theory and practice. In D. Gopher, & A. Koriat (Eds.), *Attention and performance XVII: Cognitive regulation of performance, interaction of theory and application* (pp. 223–258). International Association for the Study of Attention and Performance.
- Muradoglu, G., & Harvey, N. (2012). Behavioural finance: The role of psychological factors in financial decisions. *Review of Behavioral Finance, 4*, 68–80.
- Neal, D. T., Wood, W., & Quinn, J. M. (2006). Habits—A repeat performance. *Current Directions in Psychological Science, 15*, 198–202.
- Neal, D. T., Wood, W., Wu, M., & Kurlander, D. (2011). The pull of the past: When do habits persist despite conflict with motives? *Personality and Social Psychology Bulletin, 37*, 1428–1437.
- Payne, J. W. (1976). Task complexity and contingent processing in decision making: An information search and protocol analysis. *Organizational Behavior and Human Performance, 16*, 366–387.
- Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. New York: Cambridge University Press.
- Peterson, C. R., & Beach, L. R. (1967). Man as an intuitive statistician. *Psychological Bulletin, 68*, 29–46.
- Pickup, L., Wilson, J. R., Sharpies, S., Norris, B., Clarke, T., & Young, M. S. (2005). Fundamental examination of mental workload in the rail industry. *Theoretical Issues in Ergonomic Science, 6*, 463–482.
- Proctor, R. W., & Vu, K.-P.L. (2012). Human information processing: An overview for human-computer interaction. In J. A. Jacko (Ed.), *The human-computer interaction handbook: Fundamentals, evolving technologies, and emerging applications* (pp. 21–40) (3rd ed.). Boca Raton, FL: CRC Press.
- Rachlinski, J. J. (2000). Heuristics and biases in the courts: Ignorance or adaptation? *Oregon Law Review, 79*, 61.
- Raiffa, H. (1968). *Decision analysis: Introductory lectures on choices under uncertainty*. Oxford: Addison-Wesley.
- Ramsey, F. P. (1931). Truth and probability. In F. P. Ramsey (Ed.), *The foundations of mathematics and other logical essays* (pp. 156–198). New York: Harcourt Brace.
- Read, D., Frederick, S., & Airola, M. (2012). Four days later in Cincinnati: Longitudinal tests of hyperbolic discounting. *Acta Psychologica, 140*, 177–185.
- Reyna, V. F. (2004). How people make decisions that involve risk: A dual-processes approach. *Current Directions in Psychological Science, 13*, 60–66.
- Reyna, V. F., & Brainerd, C. J. (1991). Fuzzy-trace theory and framing effects in choice: Gist extraction, truncation, and conversion. *Journal of Behavioral Decision Making, 4*, 249–262.

- Rigas, G., & Brehmer, B. (1999). Mental processes in intelligence tests and dynamic decision making tasks. In P. Juslin, & H. Montgomery (Eds.), *Judgment and decision making: Neo-Brunswikian and process-tracing approaches* (pp. 45–65). Mahwah, NJ: Erlbaum.
- Ryan, A. M., & Ployhart, R. E. (2013). A century of selection. *Annual Review of Psychology*, 65, 20.1–20.25.
- Savage, L. J. (1954). *Foundations of statistics*. New York: Wiley.
- Schacter, D. L. (1999). The seven sins of memory: Insights from psychology and cognitive neuroscience. *American Psychologist*, 54, 182–203.
- Schmitt, N. (2012). *The Oxford handbook of personnel assessment and selection*. New York: Oxford University Press.
- Schwartz, A., & Elstein, A. S. (2009). Clinical problem solving and diagnostic decision making: A selective review of the cognitive research literature. In J. A. Knottnerus, & F. Buntinx (Eds.), *The evidence base of clinical diagnosis: Theory and methods of diagnostic research*. (pp. 237–255) (2nd ed.). Oxford: Blackwell.
- Shepperd, J. A., Klein, W. M. P., Waters, E. A., & Weinstein, N. D. (2013). Taking stock of unrealistic optimism. *Perspectives on Psychological Science*, 8, 395–411.
- Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, 69, 99–118.
- Simon, H. A. (1957). *Models of man social and rational: Mathematical essays on rational human behavior*. New York: Wiley.
- Slovic, P., Finucane, M. L., Peters, E., & MacGregor, D. G. (2002). The affect heuristic. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 397–420). New York: Cambridge University Press.
- Slovic, P., Layman, M., Kraus, N., Flynn, J., Chalmers, J., & Gesell, G. (1991). Perceived risk, stigma, and potential economic impacts of a high-level nuclear waste repository in Nevada. *Risk Analysis*, 11, 683–696.
- Smith, E. R., & DeCoster, J. (2000). Dual process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review*, 4, 108–131.
- Smith, J. F., & Kida, T. (1991). Heuristics and biases: Expertise and task realism in auditing. *Psychological Bulletin*, 109, 472–489.
- Stanovich, K. E., West, R. F., & Toplak, M. E. (2010). Individual differences as essential components of heuristics and biases research. In K. Manktelow, D. Over, & S. Elquayam (Eds.), *The science of reason: A Festschrift in honor of Jonathan St. B. T. Evans* (pp. 355–396). New York: Psychology Press.
- Steel, P. (2007). The nature of procrastination: A meta-analytic and theoretical review of quintessential self-regulatory failure. *Psychological Bulletin*, 133, 65–94.
- Stewart, T. R. (2001). The lens model equation. In K. R. Hammond, & T. R. Stewart (Eds.), *The essential Brunswik* (pp. 357–364). New York: Oxford University Press.
- Strayer, D. L., & Drews, F. A. (2007a). Attention. In F. T. Durso, R. S. Nickerson, S. T. Dumais, S. Lewandowsky, & T. J. Perfect (Eds.), *Handbook of applied cognition*. (pp. 29–54) (2nd ed.). Hoboken, NJ: Wiley.
- Strayer, D. L., & Drews, F. A. (2007b). Cell-phone-induced driver distraction. *Current Directions in Psychological Science*, 16, 128–131.
- Sweeny, K., Melnyk, D., Miller, W., & Shepperd, J. A. (2010). Information avoidance: Who, what, when, and why. *Review of General Psychology*, 14, 340–353.
- Taylor, S. E. (1980). The interface of cognitive and social psychology. In J. Harvey (Ed.), *Cognition, social behavior, and the environment* (pp. 189–211). Hillsdale, NJ: Erlbaum.

- Taylor, S. E., & Crocker, J. (1981). Schematic bases of social information processing. In E. T. Higgins, C. P. Herman, & M. P. Zanna (Eds.), *Social cognition: The Ontario symposium: Vol. 1*. (pp. 89–134). Hillsdale, NJ: Erlbaum.
- Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology*, 53, 1–25.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124–1131.
- Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211, 453–458.
- Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5, 297–323.
- Tversky, A., & Koehler, D. J. (1994). Support theory: A nonextensional representation of subjective probability. *Psychological Review*, 101, 547–567.
- Van Boven, L., Traverses, M., Westfall, J., & McClelland, G. (2013). Judgment and decision making. In D. E. Carlson (Ed.), *The Oxford handbook of social cognition* (pp. 375–401). Oxford: Oxford University Press.
- Visschers, V. H. M., Meertens, R. M., Passchier, W. W. F., & de Vries, N. N. K. (2009). Probability information in risk communication: A review of the research literature. *Risk Analysis*, 29, 267–287.
- von Neumann, J., & Morgenstern, O. (1947). *Theory of games and economic behavior* (2nd ed.). Princeton, NJ: Princeton University Press.
- von Winterfeldt, D., & Edwards, W. (1986). *Decision analysis and behavioral research*. New York: Cambridge University Press.
- Wagenaar, W. A., Keren, G., & Lichtenstein, S. (1988). Islanders and hostages: Deep and surface structures of decision problems. *Acta Psychologica*, 67, 175–189.
- Weber, E. U., & Johnson, E. J. (2009). Mindful judgment and decision making. *Annual Review of Psychology*, 60, 53–85.
- Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors*, 50, 449–455.
- Wickens, C. D., & McCarley, J. (2008). *Applied attention theory*. Boca Raton, FL: Taylor & Francis.
- Wood, W., & Neal, D. T. (2007). A new look at habits and the habit–goal interface. *Psychological Review*, 114, 843–863.
- Wood, W., Quinn, J. M., & Kashy, D. A. (2002). Habits in everyday life: Thought, emotion, and action. *Journal of Personality and Social Psychology*, 83, 1281–1297.
- Wright, G., & Whalley, P. (1983). The supra-additivity of subjective probability. In B. P. Stigum, & F. Wenstop (Eds.), *Foundations of utility and risk theory with applications* (pp. 233–244). Dordrecht, The Netherlands: Ridel.
- Wu, H.-C., Lindell, M. K., & Prater, C. S. (2013). *Process tracing analysis of hurricane information displays*. College Station, TX: Texas A&M University Hazard Reduction & Recovery Center.
- Wu, H.-C., Lindell, M. K., Prater, C. S., & Samuelson, C. D. (2013). Effects of track and threat information on judgments of hurricane strike probability. *Risk Analysis*, [Epub ahead of print].
- Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inference. *American Psychologist*, 35, 151–175.

## Chapter 19

# Experiments in Organizational Behavior

**Stefan Thau**

*INSEAD, Singapore*

**Marko Pitesa**

*University of Maryland, College Park, Maryland*

**Madan Pillutla**

*London Business School, London, United Kingdom*

## I INTRODUCTION

Organizations are social systems that pursue their goals by coordinating people to engage in joint production. Organizations are characterized by division of labor and interdependent reward structures, which result in its members trying to accomplish their own, their group's, and their organization's goals while competing against other individuals, groups, and organizations. The intense interdependence of work and the resulting mixed motives for its members lead to problems that can negatively impact the productivity of individuals and, by extension, the groups and organizations in which they work. The ultimate goal of the scientific study of *organizational behavior* is to solve problems that arise from interdependent work (cf. [Wagner & Hollenbeck, 1995](#)). In so doing, organizational scholarship aims to generate insights that can help improve the effectiveness of work (e.g., by improving coordination in organizations) as well as the experience of people at work (e.g., by improving the well-being of individuals in organizations).

Examples of interdependence problems include coordination failures, lack of trust, free-riding, and unfairness. Organizational scholarship includes examination of both the role of individual (dispositional) characteristics and features of organizational contexts on these different problems. For example, a study of unfairness in organizations may consider individual differences in social comparison orientation (i.e., individual propensity to compare their inputs and outcomes with others) or the distribution of rewards within organizational subunits to make empirical claims about the relationship between unfairness perceptions

and important work outcomes such as cooperation. Traditionally, organizational behavior scholars have primarily relied on passive observational studies rather than experiments to study these problems. A review by [Highhouse \(2009\)](#) suggests that less than 5% of the studies published in flagship management journals included experiments.

This chapter discusses unethical behaviors as an example of an interdependence-based problem that has used experiments extensively and could therefore serve as an exemplar of what can be achieved by using experiments judiciously. Unethical behaviors refer to actions that violate moral, legal, or conventional agreements. Examples include bullying or harassing co-workers, taking credit for others' work, or deceiving customers about the quality of products. The whole organization can also be a vehicle of systemic unethical behavior through more or less organized processes of workplace discrimination or exploitation. Victims of unethical behavior suffer psychologically, emotionally, physically, and economically ([Aquino & Thau, 2009](#)).

We review and discuss the study of unethical behaviors through experimental methods. The problem of unethical behavior does not define the entirety of problems relevant to organizational behavior research. However, presenting this particular problem in more detail allows us to discuss the role of the experimental method in achieving the goals of organizational behavior research in general. Our review considers the role of experiments in measurement validity and the generalizations of inferences that are common to the entire field of organizational behavior.

Traditionally, organizational behavior scholarship has primarily relied on passive observational studies, varying in the level of sophistication from cross-sectional to studies based on longitudinal data. Experiments have been rarely used ([Austin, Scherbaum, & Mahlman, 2002](#); [Dipboye, 1990](#); [Fromkin & Streufert, 1976](#); [Greenberg & Tomlinson, 2004](#); [Highhouse, 2009](#); [Podsakoff & Dalton, 1987](#); [Weick, 1965](#)). The underuse of experiments in organizational behavior is unjustified and unfortunate. Our review of the experimental study of unethical behavior in organizations suggests that important organizational problems are highly amenable to experimental study. Importantly, experimentation is the only method allowing for causal inferences and, thus, is currently the best available method to build strong and robust knowledge about causes of organizational behavior.

## II UNETHICAL BEHAVIOR IN ORGANIZATIONS

Joint work, which involves the collective engagement of a number of people, is necessary for achieving organizational goals in many situations. Joint work requires norms, rules, and conventions that promote collective welfare because it gives rise to motivational conflicts between self, group, and organizational interests ([Coleman, 1990](#); [Ullmann-Margalit, 1977](#)). However, not everyone

abides by moral and legal expectations; such violations are referred to as unethical behavior. Unethical behavior at the workplace has become a major field of inquiry in organizational behavior research due to its tremendous consequences for business (Kish-Gephart, Harrison, & Treviño, 2010). Consider the example of the negative implications for organizations of just one unethical workplace behavior, employee theft: the United States Chamber of Commerce suggests that employee theft leads to costs of as much as \$40 billion annually (U.S. Chamber of Commerce, 2013), nearly 10 times the cost of all street crime combined, including burglaries and robberies (Federal Bureau of Investigation, 2011).

Research has identified various causes of unethical behavior in organizations. These causes can be broadly classified into individual characteristics, moral issues, and organizational contexts (De Cremer, Van Dick, Tenbrunsel, Pillutla, & Murnighan, 2011; Kish-Gephart et al., 2010; Pillutla, 2011; Treviño, 1986). Individual characteristics refer to dispositional factors that make individuals more or less likely to behave unethically. Moral issue characteristics concern the nature of the (im)moral act and how it might make people more or less likely to engage in the given act. Finally, organizational context characteristics concern situational factors specific to one's organizational environment, such as explicit or implicit norms of (un)ethical behavior in the organization. We review here select experimental research that significantly advanced our understanding of each type of antecedent of unethical behavior at work, and in so doing we demonstrate why experiments are an appropriate method to investigate important organizational challenges.

First, unethical behavior is difficult to measure in a valid manner through traditional passive observational methods such as self-report measures due to the risks of socially desirable responding and self-deception (Berry, Carpenter, & Barratt, 2012; Lee, 1993). These processes can cause many individuals to underreport their true levels of unethical behaviors, leading to low variance in the variable of interest. This can make it relatively more difficult to detect relationships with important antecedent variables. Because they involve orchestrating and tightly controlling a situation, experiments allow for a reliable and valid way in which unethical behavior can be measured. Numerous experimental studies have used procedures whereby participants are given an opportunity to misrepresent their performance on a task (e.g., solving math problems or anagrams) for financial gain. Unbeknownst to participants, the experimenters design the task in a way that allows them to later compare participants' actual and self-reported performance, thus objectively measuring unethical behavior (e.g., Gino & Pierce, 2009; Pitesa, Thau, & Pillutla, 2013; Zhong, Bohns, & Gino, 2010). For example, participants are asked to solve a number of anagrams (words in which the letters are scrambled) in an allotted time frame. They are then asked to report how many anagrams they solved. However, the participants do not know that the anagrams have no solution, and the number of anagrams solved can serve as a measure of cheating (Eisenberger & Shank, 1985; Wiltermuth, 2010).

Another example of the ease of measurement of unethical behavior in the experimental setting is deception (Kern & Chugh, 2009; Murnighan, Babcock, Thompson, & Pillutla, 1993; Schweitzer, DeChurch, & Gibson, 2005). For instance, Kern and Chugh asked MBA students to negotiate with another party in a situation in which they could improve their position through deceptive behaviors. The information exchanged during the negotiation can be easily recorded and coded for whether the person behaved ethically or unethically. Although not all participants engage in unethical behavior, such approaches to measuring unethical behavior in experiments often lead to sufficient variation between experimental groups. For this reason, experiments have been used to generate numerous insights about the characteristics of individuals, moral issues, and organizational contexts as drivers of unethical behavior.

The study of individual characteristics has traditionally been conducted primarily using passive observational methods using self-report measures. These self-report studies suffer from self-selection and measurement validity issues (Zuber & Kaptein, 2013). For instance, much research on individual characteristics driving unethical behavior has examined how different moral philosophies, such as relativistic or formalistic thinking, affect people's tendency to behave unethically (e.g., Barnett, Bass, & Brown, 1994; Forsyth, 1985; Henle, Giacalone, & Jurkiewicz, 2005). This research is grounded in rationalist models of (un)ethical behavior and emphasizes deliberate analysis of moral issues and explicit formulation of evaluations of actions as guides for behavior (Haidt, 2008; Jones, 1991; Sonenshein, 2007). These studies took for granted that people's views about moral matters affect their propensity to behave unethically. Yet, recent experimental research shows that the causality might flow in the opposite direction. Own behavior may in fact cause people's judgments about the morality of a situation due to the tendency to justify one's own actions (Batson, Kobrynowicz, Dinnerstein, Kampf, & Wilson, 1997; Batson, Thompson, Seuferling, Whitney, & Strongman, 1999; Valdesolo & DeSteno, 2007, 2008). People do not want to hold negative self-views (Sedikides & Strube, 1997), so they may judge cheating as less severe as a result of their own behavior and their desire to hold a positive self-view.

Because most dispositional traits can be readily activated in the laboratory (Mischel & Shoda, 1995; Tett & Guterman, 2000), experiments can solve the self-selection and reverse causality issue by manipulating individual characteristics and studying their impact on unethical behavior. Consider, for example, individual differences in the degree to which individuals espouse ethical values. This construct has been shown to affect the propensity to behave unethically and has been measured in various ways, with the most important one being individual differences in individuals' moral identity (Aquino & Reed, 2002; Aquino, Reed, Thau, & Freeman, 2007). Moral identity refers to how central people define morality to their self-concept. The construct can be measured by traditional self-report methodology (Aquino & Reed, 2002), but it can also be readily manipulated. To manipulate people's moral identity, Aquino,

Freeman, Reed, Lim, and Felps (2009) asked participants to list either the five largest cities in the United States (the control condition) or as many of the 10 Commandments as they could. Another manipulation involves a task in which participants are led to believe researchers are interested in the analysis of their handwriting. Participants are then asked to write down a list of moral or neutral words (Aquino et al., 2007). Thinking about the moral principles associated with the Commandments or handwritten words triggered morally relevant knowledge structures in memory, activating moral identity.

A different example of research that manipulated individual factors to explain a range of unethical behaviors relevant to organizations is work on an individual power and unethical behavior (Kipnis, Castell, Gergen, & Mauch, 1976; Pitesa & Thau, 2013a, 2013b). Organizations solve problems of joint production through hierarchical arrangements, in which some people have more or less control over others' outcomes (French & Raven, 1959; Lukes, 1986; Thibaut & Kelley, 1959). These differences in outcome control translate into individual differences in people's perception of their own power. To examine experimentally whether this individual difference has a causal effect on people's propensity for unethical actions, researchers in organizational behavior can make use of robust experimental methods developed in the social psychological study of power (Inesi, 2010; Malhotra & Gino, 2011; Schweinsberg, Ku, Wang, & Pillutla, 2012; Sivanathan, Pillutla, & Murnighan, 2008). The two main power manipulations are recall primes and structural manipulations. Both have been used successfully in studying unethical behaviors. Recall primes involve asking participants to recall an episode in which they had power over another person, another person had power over them, or a neutral life episode. Structural manipulations of power involve assigning people to roles in which they either control others' outcomes or others control their outcomes. For example, participants can be given (not given) the opportunity to decide on additional rewards that other participants receive for a performed task (Kipnis et al., 1976; Pitesa & Thau, 2013a, 2013b).

We emphasize that the approach to studying the role of dispositional factors allows for strong causal inferences concerning the role of a given individual characteristic in people's propensity to behave unethically. The linkage between people's own unethical behavior and their inferences about their own dispositions is an important example demonstrating the indispensability of experimentation.

The second set of explanations for unethical behavior in organizations focuses on differences in moral issues. Much research in the area is guided by the classification of moral issue features proposed by Jones (1991), such as the magnitude of consequences (the extent to which one's actions affect other people), concentration of the effect (i.e., whether the harm is entirely concentrated on one person or distributed across many people to a proportionally smaller degree), or social consensus (i.e., whether the behavior is unambiguously seen as unethical or whether there are contrasting views as to the appropriateness

of the act). Again, moral issue features are highly amenable to experimental manipulation. Prior research used scenario methodology to vary different aspects of moral issues (e.g., [McMahon & Harvey, 2007](#); [Morris & McDonald, 1995](#)). This approach has the advantage of a high level of standardization across conditions and a tight control over the study procedure. However, research has also varied features of moral issues that participants faced in a real-world context. For instance, [Pitesa et al. \(2013\)](#) gave participants an opportunity to misrepresent their performance for financial gain and varied whether participants believed their actions would (vs. would not) affect specific other participants by manipulating the instructions of the task to suggest that overreporting of performance might harm (vs. would not harm) others (see also [Pillutla & Chen, 1999](#)). Therefore, features of moral issues can be manipulated effectively in different ways through experimentation. In addition, similar to the research on individual differences discussed previously, an additional benefit of using experimentation in this domain is that passive observational studies focusing on the role of moral issues are susceptible to the reverse causality problem. A passive observation of a correlation between people's interpretation of moral issues and their (un)ethical behavior might be because the interpretation of the situation guides behavior, but it might also be that people rationalize their actions by altering their view of the situation.

Finally, with respect to organizational context characteristics, the majority of research has used passive observational methodology to examine how the work environment affects people's propensity to behave (un)ethically (e.g., [Treviño, Butterfield, & McCabe, 1998](#); [Victor & Cullen, 1988](#); [Wimbush, Shepard, & Markham, 1997](#)). A variety of self-report scales exist that ask participants to report on their perceptions of the kind and strength of ethical values endorsed by their context. The studies pertaining to this stream of research are very informative because they allow comparability of, for example, ethical cultures across different types of organizations. At the same time, we believe it is both possible and necessary to supplement passive observational studies on the role of organizational context characteristics in unethical behavior with experimental studies. One compunction researchers might have in relation to using experimental methodology to study the effects of organizational contexts is that this factor might be difficult to reproduce in the laboratory in a manner that would lead to conclusions generalizable to actual organizations ([Greenberg & Tomlinson, 2004](#); [Griffin & Michele Kacmar, 1991](#); [Ilgen, 1985](#); [Scandura & Williams, 2000](#); [Stone-Romero, 2002](#)). However, as [Highhouse \(2009\)](#) noted, the goal of research is not to generate conclusions that generalize *to* organizations but, rather, *across* organizations. Research findings in organizational behavior that would be aimed at explaining behavior specific to the context of one particular organization would in fact arguably not be generalizable to organizations more generally. Instead, the goal of organizational scholarship is to generate insights that are generalizable across a range of possible organizational contexts. We believe that experiments allow for this kind of generalization inasmuch as they can

manipulate factors relevant to an “average” organizational context. The question then becomes whether a given study examines contextual factors that are general enough to speak to a feature common to all, or at least to a particular organizational form. Passive observational studies are typically embedded in the context of a single organization, domain, or activity. Because of this design feature, passive observations are by no means better suited to accomplish the important goal of generalization. Because experiments can distill organizational features to their fundamental and crucial elements common to many different organizations, they are particularly able to accomplish the goal of generating generalizable insights.

Several examples illustrate how the role of organizational context elements in unethical behavior at work can be tested using experimental methodology. [Aquino and Becker \(2005\)](#) manipulated organizational ethical climate, or “typical organizational practices and procedures that have ethical content” ([Victor & Cullen, 1988](#), p. 101), by explicitly informing participants of others’ expectations in relation to their (un)ethical conduct (see also [Pitesa & Thau, 2013a](#)). This manipulation varies the essence of this organizational feature, social norms ([Treviño et al., 1998](#); [Victor & Cullen, 1988](#)), and is therefore very generalizable across organizations. Similarly, [Pitesa and Thau \(2013b\)](#) examined the role of organizational accountability systems in employees’ unethical behavior. They varied whether people are held accountable for the procedure by which they arrived at a particular decision or for the outcome of their decisions. The manipulation consisted of explicitly informing participants of the way in which their behavior would be assessed, which is consistent with the way accountability expectations are communicated in the real world ([Merchant & Van der Stede, 2007](#); [Rynes, Gerhart, & Parks, 2005](#); [Tetlock, 1985](#)). At the same time, those manipulations are free from idiosyncratic characteristics of any particular organization, thus making generalization across various types of organizations using these accountability arrangements more (rather than less) likely.

Finally, another reason why experimentation is indispensable in studying the effects of organizational contexts is again the risk of reverse causality—a risk that we believe is high in the study of the interaction between individuals and their social contexts (cf. [Giddens, 1984](#)). Unethical social contexts can make people more unethical ([Cialdini, Petrova, & Goldstein, 2004](#); [Gino, Ayal, & Ariely, 2009](#); [Narayanan, Ronson, & Pillutla, 2006](#); [Pillutla & Thau, 2009](#)), but it is also the case that individual actions shape organizational contexts ([Pinto, Leana, & Pil, 2008](#)). A sufficient number of “bad apples” can create an unethical organizational context ([Pierce & Snyder, 2008](#); [Pinto et al., 2008](#)), whereas a few influential but ethical individuals can increase the ethicality of an organization ([Brown, Treviño, & Harrison, 2005](#); [Mayer, Aquino, Greenbaum, & Kuenzi, 2012](#); [Mayer, Kuenzi, Greenbaum, Bardes, & Salvador, 2009](#)). Passive observational studies can have difficulty detecting whether organizational contexts affect individuals or whether individuals affect organizational contexts.

Because experiments can systematically vary one of the two features while keeping everything else constant, they are the only means by which researchers can answer such important questions.

### III CONCLUSION

Compared to passive observation, the use of the experimental method in organizational sciences is relatively limited ([Austin et al., 2002](#); [Dipboye, 1990](#); [Fromkin & Streufert, 1976](#); [Greenberg & Tomlinson, 2004](#); [Highhouse, 2009](#); [Podsakoff & Dalton, 1987](#); [Weick, 1965](#)). In this chapter, we provided arguments and reviewed evidence suggesting this situation is unjustified and unfortunate. Our review started from the notion that the ultimate goal of the scientific study of organizational behavior is to solve problems inherent to interdependent work. We believe that to accomplish these ambitious and important goals, organizational researchers should use the most effective methods at their disposal. We focused on one key challenge of interdependent work that organizational behavior scholarship aims to solve: unethical behavior. We reviewed how experiments have been used to advance the understanding of this challenge of interdependent work and to generate practical insights for managing it. Our review shows that many key causes of unethical behaviors are highly amenable to experimental investigation. To the extent that individual characteristics and organizational contexts are also considered key explanations in other problematic phenomena in organizations, we believe the same logic for experimentation applies to those phenomena.

The case for more experimentation in organizational behavior can also be made by a broader consideration of the philosophy of science underpinning our field. As in any other positivistic science, explanations in organizational behavior research consist of causal statements and their respective boundary conditions ([Hempel & Oppenheim, 1948](#)). The empirical tests of organizational behavior explanations should therefore allow for causal inferences. An experiment with random assignment is the only methodology that makes such inferences possible. Many articles in organizational behavior using passive observational methodology discuss their inability to draw causal inferences as one of their main empirical study limitations ([Scandura & Williams, 2000](#)). Why then not use experiments to test theories in the first place? Passive observational studies, regardless of their sophistication in design or data-analytical technique, do not allow for causal inferences. All of the sophistication in design and data analysis in passive observational studies in organizational behavior (e.g., the inclusion of control variables, multiwave designs, and longitudinal designs) is brought about to approximate the causal inferences that experiments with random assignment allow for. Unless organizational behavior researchers collectively subscribe to another philosophy of science in which explanations come absent of causal statements, or they devise a methodology that allows for causal inference in the absence of random assignment to conditions and a systematic

manipulation of presumed causes, we do not see an alternative to a greater use of experiments in organizational behavior.

Another argument for an increased use of experiments is that the organizational literature heavily emphasizes the development of new theories (cf. Hambrick, 2007). Unlike some other domains of science in which a small number of theoretical paradigms guide most research in the field, organizational behavior research is characterized by a continued proliferation of theories. Top journals in the field openly require development of *new* theories for papers to be considered for publication (Hambrick, 2007; Kozlowski, 2009). Regardless of the fact that this perspective causes incentives to produce noncommensurable facts (Pillutla & Thau, 2013), the primary requirement of the tests of new theories is to provide internally consistent tests, or tests that provide conclusive evidence of causal relationships (Mook, 1983; Shadish, Cook, & Campbell, 2002). Experiments are the only approach that allows for such tests. Therefore, the insistence on new theory building in the organizational literature would suggest that much, if not most, empirical work in this domain should rely on experimentation to provide internally valid tests of the many new theories. However, as we described previously, ironically the situation is exactly the opposite: although papers are required to build new theories, the norm in the field is that most papers provide empirical tests using nonexperimental methods. If the field of organizational scholarship is serious about theory building and refinement, a drastically greater use of experimentation needs to be embraced.

One common concern about the use of experiments in organizational behavior research is that experiments most often rely on samples that are not representative for people working in organizations (Highhouse, 2009). Traditionally, most experiments have been conducted on undergraduate samples that might have no or limited work experience. It might seem inappropriate to test theories about work behavior on subjects who have no real exposure to work. We agree with the view that in certain cases, studying work-relevant phenomena might render the use of the usual subjects—undergraduate students—inappropriate. But we argue that this fact does not preclude the use of experiments for two reasons. First, many theories about how people behave at work concern more general psychological and social processes. Few of these theories make explicit assumptions about specific work contexts or experiences. It is illogical that theories making no such assumptions require empirical tests that do. For instance, research on trust, which is arguably particularly relevant to organizations (Bhattacharya, Devinney, & Pillutla, 1998; Kramer, 1999) and is heavily represented in the field of organizational behavior (Colquitt, Scott, & LePine, 2007; Derfler-Rozin, Pillutla, & Thau, 2010; Pillutla, Malhotra, & Murnighan, 2003), concerns primarily processes that readily occur in the absence of specific work contexts (e.g., perceived benevolence as a cause of trust; see Deutsch, 1960). Second, even in cases in which theories cannot be meaningfully tested using a sample that does not possess a specific work-related characteristic (e.g., experience in a certain industry) or is currently in a specific work situation (e.g., employees within an organization undergoing a major organizational change), we argue that the solution

is not to settle for (only) passive observational studies. In such cases, researchers should simply strive to conduct experiments in the field.

In emphasizing the importance of experiments throughout this chapter, we do not wish to imply that passive observational research is unimportant or should be abandoned altogether. Experiments provide internally valid tests of theories and in that way demonstrate that a certain phenomenon *can* occur, not that it *does* actually occur in the real world (Mook, 1983). In psychological science, this is often sufficient because it uncovers truths about the functioning of the human mind. In organizational behavior research, it is often not sufficient. Much organizational behavior research seeks to explain phenomena that *do* exist, such as the occurrence of more or less unethical behavior within and across organizations as a function of specific individual or contextual factors. In such cases, we advocate a mixed methods approach that includes both experimentation (allowing for an internally consistent test of the theory) and passive observation (allowing for evidence that the phenomenon of interest behaves according to the predictions of the theory in the real world).

The distinction between documenting phenomena that do exist and investigating whether and when phenomena can exist also reveals an additional advantage of using experiments in organizational behavior research. By allowing researchers to investigate what behaviors can result as a function of experimental manipulations, experimental methods can provide a way to design, modify, or change characteristics of organizations with the goal of bringing about more desirable outcomes. For instance, experiments can be used to investigate whether a specific intervention in employee training or modifications of organizational features makes employees more productive or more satisfied with their work. By merely documenting the current state of affairs through passive observational studies, it is impossible to generate this kind of useful information about the effectiveness of interventions.

In conclusion, we urge organizational scholars, particularly those testing new theories and those investigating the effectiveness of new organizational policies, to supplement their use of passive observational methods with experiments, whether in the field or the lab. This approach offers the only internally valid evidence from which researchers can draw their conclusions. A transition to this more balanced and more methodologically robust approach to organizational research is necessary if organizational science is to move closer to its ultimate goal of solving problems inherent to interdependent work and improving the effectiveness and experience of people at work.

## REFERENCES

- Aquino, K., & Becker, T. E. (2005). Lying in negotiations: How individual and situational factors influence the use of neutralization strategies. *Journal of Organizational Behavior*, 26(6), 661–679.
- Aquino, K., Freeman, D., Reed, A., Lim, V. K. G., & Felps, W. (2009). Testing a social-cognitive model of moral behavior: The interactive influence of situations and moral identity centrality. *Journal of Personality and Social Psychology*, 97(1), 123–141.

- Aquino, K., & Reed, A. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology*, 83(6), 1423–1440.
- Aquino, K., Reed, A., Thau, S., & Freeman, D. (2007). A grotesque and dark beauty: How moral identity and mechanisms of moral disengagement influence cognitive and emotional reactions to war. *Journal of Experimental Social Psychology*, 43(3), 385–392.
- Aquino, K., & Thau, S. (2009). Workplace victimization: Aggression from the target's perspective. *Annual Review of Psychology*, 60, 717–741.
- Austin, J. T., Scherbaum, C. A., & Mahlman, R. A. (2002). History of research methods in industrial organizational psychology: Measurement, design, analysis. In S. G. Rogelberg (Ed.), *Handbook of research methods in industrial and organizational psychology* (pp. 77–98). Malden, MA: Blackwell.
- Barnett, T., Bass, K., & Brown, G. (1994). Ethical ideology and ethical judgment regarding ethical issues in business. *Journal of Business Ethics*, 13(6), 469–480.
- Batson, C. D., Kobrynowicz, D., Dinnerstein, J. L., Kampf, H. C., & Wilson, A. D. (1997). In a very different voice: Unmasking moral hypocrisy. *Journal of Personality and Social Psychology*, 72(6), 1335–1348.
- Batson, C. D., Thompson, E. R., Seuferling, G., Whitney, H., & Strongman, J. A. (1999). Moral hypocrisy: Appearing moral to oneself without being so. *Journal of Personality and Social Psychology*, 77(3), 525–537.
- Berry, C. M., Carpenter, N. C., & Barratt, C. L. (2012). Do other-reports of counterproductive work behavior provide an incremental contribution over self-reports? A meta-analytic comparison. *Journal of Applied Psychology*, 97(3), 613–636.
- Bhattacharya, R., Devinney, T. M., & Pillutla, M. M. (1998). A formal model of trust based on outcomes. *Academy of Management Review*, 23(3), 459–472.
- Brown, M., Treviño, L., & Harrison, D. (2005). Ethical leadership: A social learning perspective for construct development and testing. *Organizational Behavior and Human Decision Processes*, 97(2), 117–134.
- Cialdini, R. B., Petrova, P. K., & Goldstein, N. J. (2004). The hidden costs of organizational dishonesty. *MIT Sloan Management Review*, 45(3), 67–74.
- Coleman, J. (1990). *Foundations of social theory*. Cambridge, MA: Harvard University Press.
- Colquitt, J. A., Scott, B. A., & LePine, J. A. (2007). Trust, trustworthiness, and trust propensity: A meta-analytic test of their unique relationships with risk taking and job performance. *Journal of Applied Psychology*, 92(4), 909–927.
- De Cremer, D., Van Dick, R., Tenbrunsel, A., Pillutla, M., & Murnighan, J. K. (2011). Understanding ethical behavior and decision making in management: A behavioural business ethics approach. *British Journal of Management*, 22(1), 1–4.
- Derfler-Rozin, R., Pillutla, M., & Thau, S. (2010). Social reconnection revisited: The effects of social exclusion risk on reciprocity, trust, and general risk-taking. *Organizational Behavior and Human Decision Processes*, 112(2), 140–150.
- Deutsch, M. (1960). The effect of motivational orientation upon trust and suspicion. *Human Relations*, 13(2), 123–140.
- Dipboye, R. L. (1990). Laboratory vs. field research in industrial–organizational psychology. In C. L. Cooper & I. T. Robertson (Eds.), *International review of industrial and organizational psychology* (pp. 1–34). New York: Wiley.
- Eisenberger, R., & Shank, D. M. (1985). Personal work ethic and effort training affect cheating. *Journal of Personality and Social Psychology*, 49(2), 520–528.
- Federal Bureau of Investigation. (2011). *Uniform crime reports: Property crime*. Retrieved May 30, 2013, from <http://www.fbi.gov/about-us/cjis/ucr/crime-in-the-u-s/2011/crime-in-the-u-s-2011/property-crime/property-crime>.

- Forsyth, D. R. (1985). Individual differences in information integration during moral judgment. *Journal of Personality and Social Psychology, 49*(1), 264–272.
- French, J. & Raven, B. H. (Eds.), (1959). *The bases of social power*. Ann Arbor, MI: Institute for Social Research.
- Fromkin, H. L., & Streufert, S. (1976). Laboratory experimentation. In M. D. Dunnette (Ed.), *Handbook of industrial and organizational psychology* (pp. 415–465). Chicago: Rand McNally.
- Giddens, A. (1984). *The constitution of society: Outline of the theory of structuration*. Cambridge, UK: Polity Press.
- Gino, F., Ayal, S., & Ariely, D. (2009). Contagion and differentiation in unethical behavior. *Psychological Science, 20*(3), 393–398.
- Gino, F., & Pierce, L. (2009). The abundance effect: Unethical behavior in the presence of wealth. *Organizational Behavior and Human Decision Processes, 109*(2), 142–155.
- Greenberg, J., & Tomlinson, E. C. (2004). Situated experiments in organizations: Transplanting the lab to the field. *Journal of Management, 30*(5), 703–724.
- Griffin, R., & Michele Kacmar, K. (1991). Laboratory research in management: Misconceptions and missed opportunities. *Journal of Organizational Behavior, 12*(4), 301–311.
- Haidt, J. (2008). Morality. *Perspectives on Psychological Science, 3*(1), 65–72.
- Hambrick, D. C. (2007). The field of management's devotion to theory: Too much of a good thing? *Academy of Management Journal, 50*(6), 1346–1352.
- Hempel, C. G., & Oppenheim, P. (1948). Studies in the logic of explanation. *Philosophy of Science, 15*(2), 135–175.
- Henle, C. A., Giacalone, R. A., & Jurkiewicz, C. L. (2005). The role of ethical ideology in workplace deviance. *Journal of Business Ethics, 56*(3), 219–230.
- Highhouse, S. (2009). Designing experiments that generalize. *Organizational Research Methods, 12*(3), 554–566.
- Ilgen, D. R. (1985). Laboratory research: A question of when, not if. In E. A. Locke (Ed.), *Generalizing from laboratory to field settings* (pp. 257–267). Indianapolis, IN: Heath.
- Inesi, M. (2010). Power and loss aversion. *Organizational Behavior and Human Decision Processes, 112*(1), 58–69.
- Jones, T. M. (1991). Ethical decision making by individuals in organizations: An issue-contingent model. *Academy of Management Review, 16*(2), 366–395.
- Kern, M., & Chugh, D. (2009). Bounded ethicality. *Psychological Science, 20*(3), 378–384.
- Kipnis, D., Castell, P. J., Gergen, M., & Mauch, D. (1976). Metamorphic effects of power. *Journal of Applied Psychology, 61*(2), 127–135.
- Kish-Gephart, J. J., Harrison, D. A., & Treviño, L. K. (2010). Bad apples, bad cases, and bad barrels: Meta-analytic evidence about sources of unethical decisions at work. *Journal of Applied Psychology, 95*(1), 1–31.
- Kozlowski, S. W. J. (2009). Editorial. *Journal of Applied Psychology, 94*(1), 1–4.
- Kramer, R. M. (1999). Trust and distrust in organizations: Emerging perspectives, enduring questions. *Annual Review of Psychology, 50*(1), 569–598.
- Lee, R. M. (1993). *Doing research on sensitive topics*. Thousand Oaks, CA: Sage.
- Lukes, S. (1986). *Power*. New York: New York University Press.
- Malhotra, D., & Gino, F. (2011). The pursuit of power corrupts: How investing in outside options motivates opportunism in relationships. *Administrative Science Quarterly, 56*(4), 559–592.
- Mayer, D. M., Aquino, K., Greenbaum, R. L., & Kuenzi, M. (2012). Who displays ethical leadership, and why does it matter? An examination of antecedents and consequences of ethical leadership. *Academy of Management Journal, 55*(1), 151–171.

- Mayer, D. M., Kuenzi, M., Greenbaum, R., Bardes, M., & Salvador, R. (2009). How low does ethical leadership flow? Test of a trickle-down model. *Organizational Behavior and Human Decision Processes*, 108(1), 1–13.
- McMahon, J., & Harvey, R. (2007). The effect of moral intensity on ethical judgment. *Journal of Business Ethics*, 72(4), 335–357.
- Merchant, K. A., & Van der Stede, W. A. (2007). *Management control systems: Performance measurement, evaluation, and incentives*. Essex, UK: Pearson.
- Mischel, W., & Shoda, Y. (1995). A cognitive-affective system theory of personality: Reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological Review*, 102(2), 246–268.
- Mook, D. G. (1983). In defense of external invalidity. *American Psychologist*, 38(4), 379–387.
- Morris, S., & McDonald, R. (1995). The role of moral intensity in moral judgments: An empirical investigation. *Journal of Business Ethics*, 14(9), 715–726.
- Murnighan, J. K., Babcock, L., Thompson, L., & Pillutla, M. M. (1993). The information dilemma in negotiations: Effects of experience, incentives, and integrative potential. *International Journal of Conflict Management*, 10(4), 313–339.
- Narayanan, J., Ronson, S., & Pillutla, M. M. (2006). Groups as enablers of unethical behavior: The role of cohesion on group member actions. *Research on Managing Groups and Teams*, 8, 127–147.
- Pierce, L., & Snyder, J. (2008). Ethical spillovers in firms: Evidence from vehicle emissions testing. *Management Science*, 54(11), 1891–1903.
- Pillutla, M. M. (2011). When good people do wrong: Morality, social identity, and ethical behavior. In D. De Cremer, R. van Dick, & J. K. Murnighan (Eds.), *Social psychology and organizations*. New York: Routledge.
- Pillutla, M. M., & Chen, X.-P. (1999). Social norms and cooperation in social dilemmas: The effects of context and feedback. *Organizational Behavior and Human Decision Processes*, 78(2), 81–103.
- Pillutla, M. M., Malhotra, D., & Murnighan, J. K. (2003). Attributions of trust and the calculus of reciprocity. *Journal of Experimental Social Psychology*, 39(5), 448–455.
- Pillutla, M. M., & Thau, S. (2009). Actual and potential exclusion as determinants of individuals' unethical behaviors in groups. In D. De Cremer (Ed.), *Psychological perspectives on ethical behavior and decision making* (pp. 121–133). Charlotte, NC: IAP.
- Pillutla, M. M., & Thau, S. (2013). Organizational sciences' obsession with "that's interesting!" Consequences and an alternative. *Organizational Psychology Review*, 3(2), 187–194.
- Pinto, J., Leana, C. R., & Pil, F. K. (2008). Corrupt organizations or organizations of corrupt individuals? Two types of organization-level corruption. *Academy of Management Review*, 33(3), 685–709.
- Pitesa, M., & Thau, S. (2013a). Compliant sinners, obstinate saints: How power and self-focus determine the effectiveness of social influences in ethical decision making. *Academy of Management Journal*, 56(3), 636–658.
- Pitesa, M., & Thau, S. (2013b). Masters of the universe: How power and accountability influence self-serving decisions under moral hazard. *Journal of Applied Psychology*, 98(3), 550–558.
- Pitesa, M., Thau, S., & Pillutla, M. M. (2013). Cognitive control and socially desirable behavior: The role of interpersonal impact. *Organizational Behavior and Human Decision Processes*, 122, 232–243.
- Podsakoff, P. M., & Dalton, D. R. (1987). Research methodology in organizational studies. *Journal of Management*, 13(2), 419–441.
- Rynes, S. L., Gerhart, B., & Parks, L. (2005). Personnel psychology: Performance evaluation and pay for performance. *Annual Review of Psychology*, 56, 571–600.

- Scandura, T. A., & Williams, E. A. (2000). Research methodology in management: Current practices, trends, and implications for future research. *Academy of Management Journal*, 43(6), 1248–1264.
- Schweinsberg, M., Ku, G., Wang, C. S., & Pillutla, M. M. (2012). Starting high and ending with nothing: The role of anchors and power in negotiations. *Journal of Experimental Social Psychology*, 48(1), 226–231.
- Schweitzer, M. E., DeChurch, L. A., & Gibson, D. E. (2005). Conflict frames and the use of deception: Are competitive negotiators less ethical? *Journal of Applied Social Psychology*, 35(10), 2123–2149.
- Sedikides, C., & Strube, M. J. (1997). Self-evaluation: To thine own self be good, to thine own self be sure, to thine own self be true, and to thine own self be better. *Advances in Experimental Social Psychology*, 29, 209–269.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.
- Sivanathan, N., Pillutla, M. M., & Murnighan, J. K. (2008). Power gained, power lost. *Organizational Behavior and Human Decision Processes*, 105(2), 135–146.
- Sonenshein, S. (2007). The role of construction, intuition, and justification in responding to ethical issues at work: The sensemaking-intuition model. *Academy of Management Review*, 32(4), 1022–1040.
- Stone-Romero, E. F. (2002). The relative validity and usefulness of various empirical research designs. In S. G. Rogelberg (Ed.), *Handbook of research methods in industrial and organizational psychology*. Malden, MA: Blackwell.
- Tetlock, P. E. (1985). Accountability: The neglected social context of judgment and choice. In L. L. Cummings, & B. M. Staw (Eds.), In *Research in organizational behavior Vol. 7*. (pp. 297–332). Greenwich, CT: JAI Press.
- Tett, R. P., & Guterman, H. A. (2000). Situation trait relevance, trait expression, and cross-situational consistency: Testing a principle of trait activation. *Journal of Research in Personality*, 34(4), 397–423.
- Thibaut, J. W., & Kelley, H. H. (1959). *The social psychology of groups*. New York: Wiley.
- Treviño, L. K. (1986). Ethical decision making in organizations: A person–situation interactionist model. *Academy of Management Review*, 11(3), 601–617.
- Treviño, L. K., Butterfield, K., & McCabe, D. (1998). The ethical context in organizations: Influences on employee attitudes and behaviors. *Business Ethics Quarterly*, 8(3), 447–476.
- U.S. Chamber of Commerce. (2013). *Detecting and deterring fraud in small businesses*. Retrieved May 30, 2013, from, [http://www.uschambersmallbusinessnation.com/toolkits/guide/P14\\_1000](http://www.uschambersmallbusinessnation.com/toolkits/guide/P14_1000).
- Ullmann-Margalit, E. (1977). *The emergence of norms*. Oxford, UK: Clarendon.
- Valdesolo, P., & DeSteno, D. (2007). Moral hypocrisy: Social groups and the flexibility of virtue. *Psychological Science*, 18(8), 689–690.
- Valdesolo, P., & DeSteno, D. (2008). The duality of virtue: Deconstructing the moral hypocrite. *Journal of Experimental Social Psychology*, 44(5), 1334–1338.
- Victor, B., & Cullen, J. B. (1988). The organizational bases of ethical work climates. *Administrative Science Quarterly*, 33(1), 101–125.
- Wagner, J., & Hollenbeck, J. (1995). *Management of organizational behavior* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Weick, K. E. (1965). Laboratory experimentation with organizations. In J. G. March (Ed.), *Handbook of organizations* (pp. 194–260). Chicago: Rand McNally.

- Wiltermuth, S. S. (2010). Cheating more when the spoils are split. *Organizational Behavior and Human Decision Processes*, 115(2), 157–168.
- Wimbush, J. C., Shepard, J. M., & Markham, S. E. (1997). An empirical examination of the relationship between ethical climate and ethical behavior from multiple levels of analysis. *Journal of Business Ethics*, 16(16), 1705–1716.
- Zhong, C. B., Bohns, V. K., & Gino, F. (2010). Good lamps are the best police: Darkness increases dishonesty and self-interested behavior. *Psychological Science*, 21(3), 311–314.
- Zuber, F., & Kaptein, M. (2013, October). Painting with the same brush? Surveying unethical behavior in the workplace using self-reports and observer-reports. *Journal of Business Ethics*, <http://dx.doi.org/10.1007/s10551-013-1920-y>.

## Part III

# Applied Research and Proposals

Part III contains two chapters that describe developing applied or policy implications from experiments and also practical concerns involved in writing proposals.

Chapter 20, by James E. Driskell, Jennifer King, and Tripp Driskell, describes experimental research for applied purposes, focusing on differences from (and similarities to) the more common type of experiments that are used for basic research. The authors bring decades of experience with applied research to this chapter, providing both abstract discussions and concrete instances of problems and solutions. Everything from seeking funding through design of the experiments to presenting the results may differ in applied experiments, and this chapter provides details on all the ways they differ. The authors seek to provide useful guides for researchers doing or wishing to do applied experimental research.

Chapter 21, by Murray Webster, Jr., describes the growth of external funding, with particular application to experimental research. He lists some of the sources of funding and describes the proper role of relations between funding agency officers and researchers. The chapter includes extended discussion and practical advice for writing successful research proposals.

## Chapter 20

# Conducting Applied Experimental Research

**James E. Driskell**

*Florida Maxima Corporation, Orlando, Florida*

**Jennifer King**

*U.S. Naval Research Laboratory, Washington, DC*

**Tripp Driskell**

*Institute for Simulation and Training, University of Central Florida, Orlando, Florida*

### I CONDUCTING APPLIED EXPERIMENTAL RESEARCH

The first task that we face in this chapter is to describe what we mean by the term “applied experimental research.” Previous chapters have discussed “experimental research” in considerable detail. Our goal, then, is to define what the descriptor “applied” implies as it describes a particular type of experimental research. This task is not as easy as one would hope. One way to approach this task is to refer to the journals in this field: in other words, how do the journals that publish applied experimental research describe this type of research? According to the *Journal of Applied Psychology*, applied research should “take ideas, findings, and models from basic research and use them to help solve problems that both psychologists and nonpsychologists care about” (Murphy, 2002, p. 1019). Applied experimental research should “enhance our understanding of behavior that has practical implications within particular contexts” (Zedeck, 2003, p. 3). According to the *Journal of Experimental Psychology: Applied*, applied experimental research constitutes “research with an applied orientation” that should bridge practically oriented problems and theory (Ackerman, 2002, p. 4). According to the *Journal of Applied Social Psychology*, applied research is research that has “applications to current problems of society” and that can “bridge the theoretical and applied areas of social research” (Baum, 2005, paragraph 1). Therefore, we have accomplished our first goal with at least some

success: at a broad level, applied experimental research is research that applies or extends theory to an identified real-world problem with a practical outcome in mind.<sup>1</sup>

However, this picture becomes slightly more clouded when we consider the types of content areas in which applied experimental research is conducted. Again, scanning the relevant applied research journals, we find that applied experimental research may include research on perception, attention, decision-making, learning, performance, health, race relations, group processes, leadership, work motivation, assessment, work stress, violence, poverty, legal issues, aging, gender, population, behavioral medicine, consumer behavior, sports, traffic and transportation, eyewitness memory, and other topics.

We believe that it would not be a particularly pleasant experience for the reader, or for the authors, to attempt to survey this broad domain of applications. Therefore, the objective of this chapter is somewhat less ambitious. We discuss applied experimental research in the context with which we are most familiar: applied social psychology. Our coverage of this topic is selective rather than comprehensive. Furthermore, in keeping with the title of this chapter, we intend our discussion to be practical rather than abstract. Thus, our focus is on practical issues involved in conducting applied experimental research in real-world contexts, such as the military. These issues include generating proposals and obtaining funding, conducting applied experimental research, and presenting research results to the user.

## A A Brief History of Applied Experimental Research

We can easily trace early pioneering work in applied social science back over a century, including [Triplett's \(1898\)](#) field experiments to study the “dynamo-genic factors” that affected work, [Taylor's \(1903\)](#) research for the steel industry to enhance worker productivity, Walter Dill [Scott's \(1903\)](#) research on advertising, [Thorndike's \(1903\)](#) work in education, and many others. We can further distinguish three historical periods of applied social science research. The first period is represented by these early attempts at application, in which researchers ventured out of academia to address problems of everyday life. The hope was that scientific knowledge could be applied “by enlarging its scope and making

---

1. [Cohen \(1989, pp. 53–58\)](#) distinguishes basic research, applied research, and engineering. In this categorization, basic research is oriented to producing and evaluating knowledge; applied research demonstrates the value of accepted knowledge for a practical purpose; and engineering is oriented to solving recognized problems using all available means, including well-established knowledge from basic research as well as other kinds of knowledge such as intuition and experienced-based hunches. [Webster and Whitmeyer \(2001\)](#) describe somewhat different distinctions, with examples of each type. In this chapter, we describe what Cohen would term “applied research and engineering.” As we show, these kinds of research also develop and assess new knowledge in ways similar to those of basic research.

its experiments where people work and play as well as the laboratory" (Freyd, 1926, p. 314). Moreover, applied researchers hoped that their work could replace "the lore of the folk by an array of knowledge equally concrete and practical, but immeasurably wider, more accurate, more systematic, and free from personal bias" (Bryan & Harter, 1899, p. 347; cited in van Strien, 1998).

Perhaps this pioneering spirit is most interestingly illustrated by the work of Hugo Münsterberg, who wrote one of the first texts referencing applied social science research in *Psychology General and Applied* (1914). Münsterberg was not always a proponent of applied research and in 1898 wrote a scathing attack on then current attempts to apply psychology to the educational system, claiming that "our laboratory work cannot teach you anything which is of direct use to you in your work as teachers" (Münsterberg, 1898, p. 166). This statement is characteristic of Münsterberg because he seemed to be a polarizing individual in general. Benjamin (2006) has noted that although Münsterberg was one of the leading applied psychologists of the time, he was also "one of the most despised individuals in America" (p. 414), due partly to his personal egotism and pomposity but primarily to his prominence as a leading proponent of German involvement in World War I. In any case, Münsterberg later reversed his position on the value of applied research, and in fact he believed so strongly in the importance of applied research that he suggested that the government establish experimental research stations throughout the country, similar to the agricultural extension stations that were being introduced at that time. Although agricultural extension stations were established in every county or district in every state, the case for applied social science research stations was probably a bit more difficult to sell.

A second historical period of applied social science research is represented by a measured backlash or at least professional ambivalence regarding applied research. The early attempts to apply principles of social science to the workplace were admittedly "seat-of-the-pants" efforts, and as van Strien (1998) noted, "'Bona fide' psychologists, as academics saw themselves, looked on the crude methods of nonacademic practitioners with great reservation" (p. 220). Again, reference to Munsterberg is illustrative of the ambivalence regarding applied research during this period, as he both conducted applied research and also railed against applied research that lacked strong scientific foundation. Yet as Benjamin (2006) noted, Münsterberg himself rarely abstained from commenting (or publishing) on almost any topic whatsoever, and he was broadly criticized by other scientists for forsaking scientific standards for the sensationalism and faddism of popular pseudoscience.

Moreover, during this period, academic scientists were particularly concerned with distinguishing themselves from the phrenologists, psychics, mediums, palm readers, and spiritualists that had captured the public imagination. Therefore, although applied scientists were to attack the problems of the day with some success, there was a stigma attached to applied research and an ambivalence regarding those who would forsake the academic corridors for the

outside world that continues to this day. Although we do not wish to overstate the case, there exists in many instances a bias toward a “pure” academic culture in the social sciences that views applied research as suspect.

A third historical period paralleled the applied research activities that were initiated during World War I and conducted intensively during World War II. It would not be unfair to suggest that the “person-in-the-street” in the early part of the 20th century viewed the social science enterprise with at least some degree of indifference. In fact, when social scientists first approached the U.S. War Department in 1917 to propose research to support the war effort, the response was initially lukewarm ([Driskell & Olmstead, 1989](#)). However, the applied research that was conducted during World War I in areas such as selection, training, and motivation constituted the first large-scale effort in the social sciences to apply principles derived from the academic laboratory to address society’s needs. In fact, the military provided a particularly fertile ground for the examination and application of scientific theories.

Thus, psychologists were able to study learning, human performance, and morale; sociologists could study conflict, race relations, and the military organization; and political scientists could study international relations and policy. One research project called for the assessment and testing of more than 4,000 recruits arriving at a military camp. The statistical team for this project included such luminaries as E. L. Thorndike, L. L. Thurstone, and A. S. Otis, and the cost to the War Department for this work was less than \$2,500 ([Driskell & Olmstead, 1989](#)). As the nation faced World War II, military program managers were able to draw on the relations formed with the scientific community in World War I, and the subsequent applied research conducted during World War II had a tremendous impact in both: (1) contributing to the war effort; and (2) legitimating and demonstrating the value of applied social science research.

This success is best illustrated by reference to a classic wartime research program that became known as *The American Soldier* research, summarized in a four-volume series titled *Studies in Social Psychology in World War II* (see [Stouffer et al., 1949](#)). Contributors to this research included Samuel Stouffer, Robin Williams, Robert K. Merton, Edward Shils, Paul Lazarsfeld, Brewster Smith, Rensis Likert, Arthur Lumsdaine, Irving Janis, Carl Hovland, Quinn McNemar, and Louis Guttman. Stouffer et al. summarized this work as a “mine of data, perhaps unparalleled in magnitude in the history of any single research enterprise in social psychology” (p. 29). Certainly, the broader contributions to social science made by these leading scholars was as impressive, including the work of Stouffer and Williams on relative deprivation; McNemar, Guttman, and Likert on scaling; Lazarsfeld on quantitative methods; and Merton (with Alice Rossi) on reference groups. In brief, the social and behavioral sciences stepped up to the plate in World War II and hit a home run. The successful application of social science knowledge in addressing practical wartime needs during World War II not only established this field in the public eye as having

considerable practical value but also contributed significantly to the subsequent postwar growth and development of the field as a whole.

## II BASIC AND APPLIED RESEARCH

Although we attempted to neatly define applied experimental research in the opening paragraphs of this chapter (and declared some measure of success in doing so), in practice we find that this type of work is not so easily categorized. In the following, we discuss the distinction between basic and applied research and then briefly address the role of theory in applied research.

To address the topic of basic versus applied research, we first turn to a seminal figure in the development of science policy and the creation of a scientific infrastructure in the United States, Vannevar Bush. To briefly summarize an illustrious career, Bush was an MIT engineer who became president of the Carnegie Institution (now Carnegie Mellon University) and then director of the federal Office of Scientific Research and Development during World War II. After considerable success leading the wartime research efforts, Bush was approached by President Roosevelt at the conclusion of the war with the request to provide a recommendation on how the lessons learned and successes achieved could be continued in peacetime. Bush's response, summarized in *Science: The Endless Frontier* (Bush, 1945), became the blueprint for the establishment of the National Science Foundation (NSF). In this document, Bush wrote:

*Basic research is performed without thought of practical needs. It results in general knowledge and understanding of nature and its laws. The general knowledge provides the means of answering a large number of important practical problems, though it may not give a complete specific answer to any one of them. The function of applied research is to provide such complete answers.... Basic research leads to new knowledge. It provides scientific capital. It creates the fund from which the practical applications of knowledge must be drawn. (Bush, 1945, “The Importance of Basic Research” section, paragraph 1)*

This is an exceptionally eloquent statement, but as is often the case, its elegance obscures some further qualifications regarding the distinction between basic and applied research. First, the terms *basic research* and *applied research* refer to ideal types. In its purest form, basic research involves the testing of theory for the purpose of understanding fundamental processes. The benefits of basic research are long-term, and in many cases, the societal payoff may be in areas that were not even envisioned by the original basic researchers. In its purest form, applied research involves applying theory to identified real-world problems. The benefits of applied research are short-term, and the results have an immediate and identified use or application. However, there is a considerable amount of research that is carried out between these two poles. Rarely are basic researchers not at least cognizant of the broader practical implications of their

work, and rarely do applied researchers conduct their work without concern for how their results may extend or elaborate theory.

Moreover, there is no one criterion that separates basic from applied research. Some have noted that the motivation of the researcher is a primary factor. Basic researchers will view the primary value of their work as building theory or expanding a body of knowledge, whereas applied researchers will view the primary value of their work as solving a real-world problem. Others have argued that a primary factor separating basic from applied research is the temporal nature of the contribution of the research. The practical payoff from basic research may be years away or simply may not be seen as a significant or relevant question, whereas the practical payoff from applied research is its *raison d'être*, and the results are intended to be put into use in the short term in a specific context.

It is further useful to note that even the most basic research should have identifiable societal implications, especially from a political standpoint. For example, NSF funds basic research, so it is the primary funding source for basic research in the social sciences, although other agencies sometimes also fund basic research. Although NSF focuses on basic research at the frontiers of knowledge, some NSF-funded results are immediately useful, and NSF often goes to great length to tout these as success studies. Emphasizing these "discoveries" is one way to demonstrate to legislators and other interested parties that the national investment in research and development is being put to good use. The fact that research, both basic and applied, must meet some foreseeable national good recalls a famous anecdote: when the British physicist Faraday was asked by the Finance Minister Gladstone in the 1850s whether electricity had any practical value, Faraday replied, "One day Sir you may tax it." It is quite likely that this reply resulted in a well-funded program of research as well as a delighted politician.

In one of our favorite quotes, [Melton \(1952\)](#) facetiously labeled basic research as "what I want to do" and applied research as "what someone else wants me to do" (p. 134). Although this may not be as elegant a proclamation as that provided by Vannevar Bush, there is certainly an element of truth in this statement. In practical terms, there is a clear and broad distinction between research designed to test theory and research designed to apply theory. Moreover, as we note in a subsequent section of this chapter, research designed to test theory and research designed to apply theory are conducted in a different manner. Nevertheless, basic and applied research go hand-in-hand, with applied research serving as a bridge between basic research and real-world applications.

### **III THE ROLE OF THEORY IN APPLIED EXPERIMENTAL RESEARCH**

Popular culture draws a sharp distinction between theory and application, depicting theorists as slightly befuddled intellectuals and applied researchers as

unprincipled mercenaries. To counter this perspective, we offer psychologist [Kurt Lewin's \(1944/1951\)](#) famous quote regarding theory and application:

*Many psychologists working in an applied field are keenly aware of the need for close cooperation between theoretical and applied psychology. This can be accomplished in psychology, as it has been accomplished in physics, if the theorist does not look toward applied problems with highbrow aversion or with a fear of social problems and if the applied psychologist realizes that there is nothing so practical as a good theory. (p. 169)*

This statement is most informative in the present context in that it evokes the close interplay between theory and application. As Vannevar Bush so pleasantly stated, basic research and theory provide the fund or capital from which practical applications can be drawn. Joseph [Berger \(1988\)](#) noted that a theory is grounded in its applications and that these applications further shape the theory in its development. Turner (1998, p. 247) warns against theories “undisciplined by empirical research” and notes that theories that are tested and used on real-world problems not only provide useful information about the problems of the world but also test the utility of the theories themselves. Thus, the relationship between theory and application is reciprocal.

On the one hand, applied research takes the concepts and principles from basic theory and uses them to solve real-world problems. Thus, applied research is guided by theory, and applied research that is devoid of theory is itself indistinguishable from the work of the charlatan or seer. On the other hand, theory can grow and develop because the results of applied research often present “problems” that the current theory cannot explain. Moreover, one criterion that determines a theory’s value is the extent to which it can be applied usefully in various settings. Thus, in this sense, theory both guides and follows applied research, and applied research both guides and follows theory.

## IV DEVELOPING A PROPOSAL AND GENERATING FUNDING FOR APPLIED EXPERIMENTAL RESEARCH

Given our interests in addressing practical aspects of conducting applied experimental research, it does not get any more practical than asking, “Who funds applied research in the social sciences?” Luckily, the answer is readily available in the annual compendium titled “Federal Funds for Research and Development” published by the [National Science Foundation \(2013\)](#). Summary data captured in more than 350 pages of charts and figures reveal that the overall level of federal support for applied research in the behavioral and social sciences in FY 2012 was approximately \$1.8 billion. (Interestingly, basic research funding in these fields for 2012 was approximately \$1.4 billion, revealing a seemingly reasonable balance, give or take a couple hundred million dollars.)

Federal funding categories are separate for psychology and the social sciences, so we summarize these data separately. Within the field of psychology,

funding for applied research comes primarily from the Department of Health and Human Services (providing approximately 83% of applied research funding, approximately \$797 million) and the Department of Defense (5%, \$52 million). Other major contributors include the Department of Transportation (4%, \$37 million) and the Department of Veterans Affairs (4%, \$42 million).

Within the social sciences, the largest funding sources of applied research include the Department of Education (22%, \$198 million), Department of Health and Human Services (21%, \$187 million), and the Department of Agriculture (16%, \$141 million). Other major contributors include the Social Security Administration (10%, \$94 million) and the Department of Commerce (8%, \$74 million).

However, those numbers indicate general funding levels for applied research and do not necessarily reflect the extent or sources of funding for applied *experimental* research in psychology and the social sciences. There are no existing data of which we are aware that would provide this more specific information, so we again adopt the strategy of perusing relevant applied experimental research journals to scan authors' acknowledgments of funding support. It seems that a considerable amount of funding for applied experimental research comes from federal agencies such as the Department of Health and Human Services, Department of Defense, and the NSF, along with a smattering of private funding sources. It may be surprising to see NSF as a significant funding source for *applied* experimental research given NSF's primary basic research mission. However, [Gerstein, Luce, Smelser, and Sperlich \(1988\)](#) noted that whether research is termed basic or applied may have more to do with the funding agency's stated mission rather than some intrinsic aspect of the research being supported. In other words, even with a basic research mission, it is not surprising that NSF in practice supports a good bit of theoretically focused empirical research that has an applied orientation.

To a considerable extent, seeking funding for "theoretically focused empirical research with an applied orientation" from NSF does not differ from seeking funding for "theoretically focused basic research" from NSF. (Webster, in Chapter 21 of this volume, discusses seeking funding for basic research.) Therefore, we focus the following discussion on funding for applied experimental research from the Department of Defense (DOD), for two reasons. First, the DOD provides approximately \$74 million in funding for applied research in the social and behavioral sciences annually ([National Science Foundation, 2013](#)), a substantial amount by any standard. Second, the authors have had considerable experience with applied DOD research, and one author (J. Driskell) has served on both sides of the fence—as an applied researcher and recipient of multiple research contracts as well as a government researcher/program manager responsible for funding such research.

It must first be considered that the DOD is an applied organization. Although research performed for the military is often indistinguishable from that performed in the university experimental laboratory, military research has a special

characteristic: it is always subject to “audit”—that audit being “How does this contribute to improved military operations?” Research is evaluated by military sponsors in terms of how well it will contribute to solving operational needs. In other words, most military research starts with a military requirement, and the end product of most research is an application (near or long-term) to an identified military problem.

In seeking DOD research funding, you may deal with two types of organizations: *funding offices* (e.g., the Office of Naval Research) and *performing organizations* (various DOD research laboratories). The funding offices, most of which are located in the Washington, DC, area, fund research directly. Typically, a Broad Agency Announcement (BAA) will be issued periodically that describes the types of research that the organization is interested in funding, and research proposals are submitted according to the instructions therein. This approach represents a “formal” proposal mechanism, and the procedure of proposal preparation and submission is similar to that of other federal agencies, including NSF.

The performing organizations, which include various government laboratories and research centers scattered throughout the country, receive their funding from the funding offices to support both in-house research and external research. These organizations may also issue announcements describing their research interests, but often the procedure of proposal preparation and submission is more informal. For example, it is likely that an astute researcher seeking funding may scan research reports in an online database such as Defense Technical Information Center’s Scientific and Technical Information Network (STINET) service (<http://stinet.dtic.mil>), find a government laboratory researcher doing related research, make contact, find out that researcher’s interests and needs, and develop funded research out of that contact.

It may be worthwhile, in seeking funding, to know thy target. Typically, the government laboratory researcher/program manager is a Ph.D.-level scientist who has gone directly from graduate school to employment in the research organization and who has extensive in-depth knowledge of his or her relevant applied content area. The government researcher usually serves in a dual role, serving as an in-house researcher as well as a program manager with budget and funding authority. For an applied scientist seeking research funding, the government program manager can serve as your person “on the inside,” familiar with the requirements of his or her organization and current research needs. From the perspective of the government program manager, the applied scientist represents someone who can help solve a problem.

As noted previously, there is considerable money invested in research each year, and the money draws the attention of some who are more interested in getting the dollars than in using them to conduct high-quality research. Government researchers are usually alert to such individuals. From experience, we note that this government researcher often encounters several types of persons seeking research funding. The first is the *businessperson/generalist*. This person may

represent a private company (often staffed largely with master's-level personnel) that can perform a variety of studies or analyses for the government organization, although they would typically not get involved in experimental research. The businessperson/generalist is a jack of all trades and master of none, but he or she can provide resources and personnel to perform a number of lower-level research tasks.

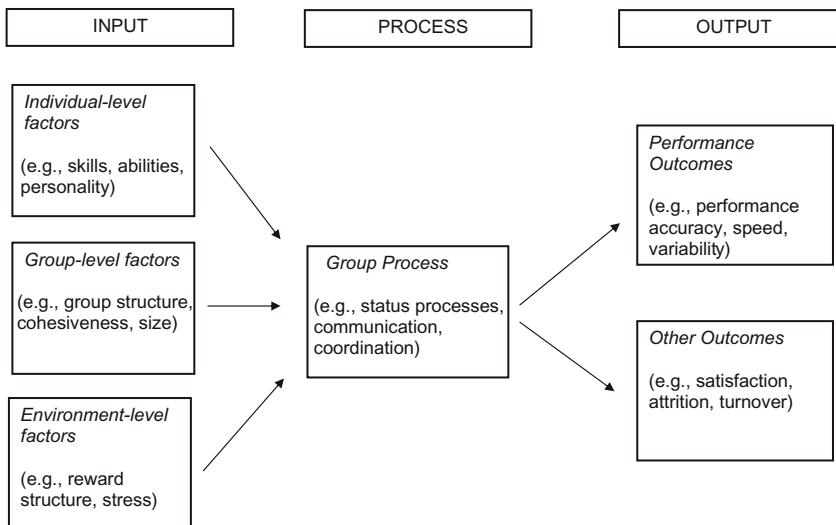
The second type is the *academic purist*. This represents an ivory tower scientist with the mind-set of "How can you fund my work?" rather than "How can I apply my expertise to address your needs or interests?" In this case, the academician is interested in doing his or her own work without oversight but wants the applied government scientist to fund it. This often does not go over very well.

A third type of person is the one who the government researcher hopes to find and often does: the *applied scientist*. The applied scientist offers expertise in a specific content area, is interested in finding out the interests and needs of the government researcher and the agency's mission, and is flexible and adaptive in adapting his or her expertise to address these requirements.

Certainly these types are caricatures, but they represent three different ways of interacting with the potential research sponsor: (1) "I can do anything"; (2) I'll do what I want to do, but I want you to pay for it"; or (3) "I have an ongoing program of research that I believe will be of interest to you, but first tell me about your research needs." It may be worthwhile for the applied researcher to consider the manner in which he or she chooses to approach a potential applied funding source. It is our experience that government sponsors welcome well-intentioned applied researchers who are willing to listen to the sponsor's needs and apply their expertise to address these problems.

In discussing the informal interaction that often occurs between the research sponsor and the applied researcher when discussing potential research, we further note that the applied researcher must be flexible in adapting his or her research interests to the requirements and interests of the sponsor. For illustration, let us consider the field of group dynamics, broadly defined. Group researchers have often drawn on an input–process–output model, as shown in [Figure 20.1](#), to organize the variables that are relevant to group interaction (see [Hackman & Morris, 1975](#)). A basic researcher interested in group status processes, for example, may conduct research with a goal of developing a better fundamental understanding of status processes. However, an applied research sponsor will, almost by definition, be interested in performance outcomes also, as represented in the rightmost column of [Figure 20.1](#).

Let us say that an applied researcher is interested in, for example, the role of status in determining challenging and monitoring behaviors in an aviation aircrew (referring to the difficulties first officers (copilots) often experience in issuing challenges or warnings to captains; see [Milanovich, Driskell, Stout, & Salas, 1998](#)). However, the research sponsor who may be interested in this topic will also be interested in two further questions. The first question is how



**FIGURE 20.1** An input–process–output model of group interaction.

this phenomenon affects performance outcomes. An applied research sponsor interested in studying groups is ultimately going to be concerned with group outcomes, such as how accurately groups perform or how quickly groups perform. A second question is how this phenomenon is affected by other contextual factors that are important in the real-world setting of interest, such as the type of group examined, the type of task undertaken, or how effects are sustained over time. It is the nature of applied research that any single phenomenon must be examined in the broader context in which it occurs. Our intent in this discussion is not to discourage anyone from entering the challenging world of applied experimental research but, instead, to note that the singular focus that serves the basic researcher so well may have to be widened as the applied researcher examines a specific phenomenon in context in the real-world.

You are likely to hear the potential research sponsor describe research programs in numerical terms, as being 6.1 (six one), 6.2 (six two), or 6.3 (six three) research. Federal research and development is divided into separate activities denoted by account numbers. The types most likely to be encountered by the applied social scientist include basic research (6.1), applied research (6.2), and advanced technology development (6.3). The basic research (6.1) category consists of scientific studies that develop the knowledge base for subsequent research or application. The applied research (6.2) category consists of R&D oriented toward a specific military problem. It includes the exploration of new technologies or concepts that hold promise for application to specific military needs. The advanced technology development (6.3) category consists of advanced development and feasibility demonstrations in operational settings.

It is important to note that the term *basic research* has a somewhat different interpretation in an applied organization such as the DOD than the term has in an academic environment. Even basic research is undertaken with an applied orientation and with a goal of transition to 6.2 and 6.3 R&D and toward later application. In other words, the military does not fund research to enhance the state of the art; it funds research to enhance the state of the military.

## A The Proposal

Although the proposer will ideally have a general understanding of the funding organization's program needs or requirements, it is not entirely clear or explicit what criteria government program managers use on an informal basis when evaluating potential new research efforts. Certainly, the proposed research must be scientifically sound, it must be relevant to some identified need and capable of being defended to others in the organization and in higher offices, and it must have a payoff or product that is seen as valuable. If the proposer is successful in making this case during preliminary discussions with the government sponsor, the proposer may be invited to submit a white paper. A white paper is a short, 3- or 4-page synopsis of a research idea. A white paper also represents a decision point: if a white paper captures the attention of the government program manager and other decision-makers in the organization, the proposer may be asked to submit a more detailed or full proposal. If not, the white paper ends up in one of the government's many file drawers. Therefore, it is incumbent on the applied scientist to be able to capture and present complex ideas in a nutshell, which is a difficult task for many of us who are more accustomed to presenting complex ideas in an elaborate and comprehensive manner.

Another document that is often requested with the submission of a white paper is a "quad chart," which is a single PowerPoint slide or page partitioned into four quarters. Each quadrant has a heading, and the quadrant headings may include: (1) title and proposer; (2) technical approach; (3) operational requirement or need addressed; and (4) research cost and schedule. Each quadrant may have room for three or four short sentences. If you think it is difficult to put a complex research proposal into four pages, try putting it into four sentences! At any rate, a quad chart may be the only thing a potential decision-maker scans, so it also represents a decision point by which a potential proposal can be weeded out or advance.

Based on our experience, we would hazard a guess that a mid-level program manager evaluating multiple potential new research starts may consider the operational need first—does the quad chart or white paper identify an operational need that is of current importance?—and then consider the technical approach or research proposed to meet that need. Typically, when developing a white paper, the research proposer can rely on the government sponsor's knowledge of current requirements to provide input to ensure that the white paper addresses

these issues. Again, it is entirely appropriate, and certainly in the researcher's interests, to contact the government program manager to discuss these and other programmatic issues.

It is also the case that some research topics are at any point in time more glamorous than others, or are hot topics in the military R&D community, and a proposal that at least references these topics may be more successful than one that does not. It might be a stretch to tie, for example, proposed research on group status processes into current interests in "network-centric operations," but it might be advantageous to do so if possible. At the least, the appropriate use of relevant buzz words may indicate to the reader or evaluator that the proposer is cognizant of current interests and requirements. We do not intend to be cynical in making this suggestion—the applied researcher is not expected to be a military subject-matter expert but, with a little bit of background research, should be knowledgeable enough to "know the lingo" and be familiar with the general research agenda of the funding organization.

A further consideration in developing a white paper or proposal is the identification of a potential user. Applied research is research that is intended to be put to use. Ideally, each research project will have a near-term user identified, even to the extent of having someone in authority in the field available who can pick up the phone and say, "Yes. I can put this research into use." Regrettably, this is not often the case in the social sciences. Compared to a scientist who is developing a new computer chip, our research rarely gets the end user leaping out of his or her seat. Although the ideal is to have some identified "transition" for research results in place, this is often not the case for experimental research, and it may be more the norm for advanced development efforts that are ready to be fielded. For applied experimental research, it is usually sufficient that the researcher and sponsor simply identify *potential* users or distinguish clear *implications* of one's research to establish a foundation for later transition of research results.

If your proposal is selected for award, you will be issued a contract or grant. In terms of how the research will be carried out, there is little distinction between whether research is conducted via a contract or a grant. There are, however, federal regulations that distinguish between the two. In general, a grant is issued by a federal agency to carry out a public purpose or stimulate a specific desired activity. A contract is issued by a federal agency to acquire goods and services for its direct benefit and use. Therefore, basic research is most often carried out via a grant, and applied research is often procured via a contract.

In practice, there are some distinctions between the contract and grant mechanisms. Contracts are essentially procurement instruments for obtaining goods or services, so they include a specific statement of work describing the services required and also include a set of milestones and deliverables for accomplishing this work. Grants are financial support for a public purpose; the statement of work is more loosely defined, and the milestones and deliverables are less explicit. In most cases, an applied researcher would be delighted to receive either

a grant or a contract to support proposed research. At any rate, whether a grant or contract is awarded is determined by the regulations of the funding agency and is largely non-negotiable.

## V CONDUCTING APPLIED EXPERIMENTAL RESEARCH

Unlike basic research, which is essentially a hands-off endeavor from the point of view of the research sponsor, applied research implies some type of relationship with the sponsor. At the least, applied research implies a reciprocal relationship between the applied researcher, who offers expertise, and the research sponsor, who offers an opportunity or requirement for application. In practice, this relationship can range from benign neglect to a desire on the part of the sponsor to be actively involved in the research. However, the research sponsor's involvement often takes the middle ground: it is often the case that the government program manager who is sponsoring and funding a particular project has shared interests with the applied researcher but is too busy managing multiple projects to become heavily involved in any particular experimental study. Nevertheless, in applied research, the government sponsor is often a valuable resource in securing entrée to research sites, obtaining required permissions, and ensuring research participation.

In most cases, applied experimental research consists of research conducted in an experimental laboratory setting for applied purposes. Many have argued that the primary purpose of the experimental laboratory is to test theory (see Driskell & Salas, 1992; Mook, 1983; Webster & Kervin, 1971; Webster & Sell, Chapter 1, this volume). Moreover, one defining characteristic of the experimental laboratory, its artificiality, is advantageous in that it allows the researcher to create controlled conditions most conducive to testing theory. That is, the greater the artificiality of the experimental setting in terms of isolating only those variables relevant to the hypothesis being tested, the greater confidence we have that we have provided a clear test of the theory.

Experimental research to test theory, as we noted previously, is a common type of basic research. However, applied experimental research often has as its goal *applying theory* to a specific setting. To the extent that basic research designed to test theory seeks control over actors, tasks, and context (see Ilgen, 1986), applied research seeks to apply theory in situations in which we are concerned with specific actors, tasks, or contexts. Berger (1988) noted that this type of applied research can serve to instantiate theory, establishing the extent of its application in various real-world contexts. Therefore, applied experimental research is often conducted to apply theory to address real-world problems by incorporating actors, tasks, or contexts specific to that setting. This type of applied research can be conducted: (1) in academic laboratories, often incorporating realistic simulations or task abstractions; (2) in an experimental setting in the target or user community; or (3) in the field utilizing the actual tasks that are performed on a daily basis.

These three types of applied research, in order, represent an increasing level of realism in terms of actors, tasks, and context reflecting the real-word environment of interest, as well as a decreasing level of control in experimentation. Howell (1998) noted that the general research strategy of conducting applied experimental research using realistic simulations or abstractions of real-world tasks can serve as a bridge between laboratory experimentation on the “basic” side and field research on the “applied” side.

The following describes one example of this type of applied experimental research conducted in the user community. Driskell and Radtke (2003) conducted empirical research to examine the effect of gesture on speech production and comprehension. Previous researchers had hypothesized that conversational hand gestures improve communication by enhancing speech, conveying information that augments the information provided by the speech channel (Beattie & Shovelton, 1999; Kendon, 1983). However, others had hypothesized that the primary function of gesture is to assist the speaker in formulating speech by aiding the retrieval of elusive words from lexical memory (Krauss, 1998; Krauss, Chen, & Chawla, 1996). Moreover, these researchers argued that speech production plays a mediating role in the observed relationship between gesture and comprehension. That is, those studies finding that gestures enhance communication may have observed this result simply because they did not control for the possibility that speakers who are allowed to gesture produce more effective speech. From a practical standpoint, the research sponsor was interested in computer-mediated communication and the question of whether one should be concerned with designing a video interface to display gestural information in computer-mediated communication when some argue that gesture provides little communicative information in the first place.

We designed an experiment as a basic analog of technical communications to represent a setting in which the speaker knows something and is trying to convey it to the listener. We conducted the experiment at the sponsor’s site using on-site personnel as research participants, varied the extent to which participants were allowed to gesture on the laboratory task, and developed measures of speech production and listener comprehension. The results indicated that gestures enhanced both listener comprehension and speech production, and that gestures had a direct effect on listener comprehension independent of the effects gestures had on speech production. Furthermore, the research implied that the ideal visual field for a computer-mediated communication system should include information communicated via gestures. Although this research addressed a very specific question, it has implications for a variety of instances in which the scope conditions are the same. For example, in both the *Columbia* and the *Challenger* disasters, communications problems were central, and these problems concerned the lack of video connections that could have ensured more effective communication (see Vaughan, 1996; Langewiesche, 2003).

The purpose of this research was to apply hypotheses derived from communications theory to a military computer-mediated communications environment.

We attempted to incorporate greater realism in this application in terms of the actors (using military personnel as research participants), the task (developing an approximation of the real-world task in the laboratory), and the context (defining the experimental situation in terms of an operational setting that the participants would care about). We use this example as an illustration of applied experimental research conducted in the user community and to set the stage for addressing some of the unique practical problems that can occur in this setting. Applied experimental research that is conducted at a user or real-world site differs in several respects from typical experimental research. We now discuss several of these concerns.

## A Recruiting Research Participants

The first required task is to find a user community that is willing to participate in the research. Sometimes this is already established and a user community is directly involved at the initiation of the research program. However, as noted previously, often an applied research program does not have an established user community at its beck and call, but it is conducted because it has clear implications for application. In this case, the government sponsor may assist in locating a community that will cooperate by providing participants and a setting for the research. For the DOD, this is typically a military community or field organization that is convenient to the researchers and that will participate out of a sense of organizational support, although it may not typically receive any direct or immediate payoff from the research itself. Research is conducted at the host organization on a not-to-interfere basis. This means that the researcher will often receive an unused classroom or offices for laboratory space and can request that research participants be scheduled at desired times, although this schedule must not interfere with the day-to-day operations of the organization (which may mean scheduling on evenings and weekends).

Typically, the researcher is assigned a point-of-contact person whose job is to ensure research participants are available according to the research schedule. In some cases, participants may not show up because, on any given day, organizational requirements or unexpected “fire drills” may supersede the research schedule. This must simply be accepted. In some cases, such “no-shows” can occur because the point of contact is too busy to track participants. This must be dealt with diplomatically. Often, scheduling can go *too* precisely—the point of contact may send enough participants for an entire day’s schedule, all showing up at 8:00 a.m. and instructed to wait their turn. In this case, the researcher must do the best he or she can to ensure that the participants are treated with appropriate care and respect.

## B Laboratory Procedures

In contrast to the relative homogeneity of an undergraduate student research pool, research populations in real-world organizations can vary considerably. In the military, the participant population may range from new recruits with a

minimum of education or experience to high-level officers with advanced education and training. Care must be taken to develop laboratory procedures that are appropriate for the participants. Furthermore, because of the heterogeneity of the population, it is especially important to ensure random assignment of participants to experimental conditions. Unless precautions are taken, it may be likely that one week, participants will come from a mechanical division, the next week from a military police division, and so on.

In applied research, it is often easier to garner research participants' interest and motivation because you are using a research task and an environment that is important to them. Therefore, it is important to establish clearly in the experimental instructions not only what you are doing and why you are doing it but also why this research is important for the military. In general, we have found that military research participants are genuinely interested in research that is being done in "their" world, and they are quite motivated to take part. On the other hand, one problem with using a research task that often only approximates the real-world task of interest is that the participants know from experience that this is not the exact task that they do in the real world, and they will let you know this in no uncertain terms.

Finally, in debriefing participants, the same theme of why the research is valuable to them and the organization should be emphasized. The research participants are part of a larger organization, and especially in the case of the military, it is an organization to which there is a high level of individual commitment. Thus, participants are particularly interested in the value of the research for the military. This emphasis on relevance can help avoid *contamination* of the research population, which refers to individual subjects telling others about critical features of the experimental situation, rendering these others unfit for participation in the study. Contamination of future research participants is less likely to be an issue if debriefing procedures are comprehensive and emphasize why the organization would be harmed if future participants are told of experimental details in advance.

## C Ethical Concerns

Ethical principles require that researchers obtain the informed consent of research participants. Informed consent includes the right to decline to participate in a research study. In the case of applied research conducted for the military, the host organization is a federal agency with its own regulations, which include the authority to have military personnel participate in research activities. In other words, research participants are generally "volunteered" by superiors to take part in a given research study. Nevertheless, the applied researcher is not relieved of ethical responsibility to ensure the welfare of research participants. As Haverkamp (2005) noted, the researcher has an obligation to promote the best interests of the participant, and this obligation cannot be set aside by the participant's consent or, in this case, by the organization's authority.

Therefore, it is good practice to obtain informed consent from research participants regardless of whether their participation is mandated by the organization. Given that the individual has been directed to participate in this activity, the researcher does not have the authority to release him or her from this obligation. What we have done in practice is to provide an alternative for those who wish to decline from participating in a study, such as allowing them to sit in the waiting room during the scheduled period and then returning to duty as previously instructed. This helps to achieve two goals. First, those who do not wish to participate can decline and still meet the requirements of their organization. Second, it ensures that those who do participate are more likely to meet basic scope conditions of the research study, such as caring about the outcome of the task.

## VI PRESENTING RESEARCH RESULTS

When doing basic research, publishing the results of one's work is part of the scientific enterprise. When doing applied research, the motivation to publish can be slightly more prosaic. We noted previously that applied research is research meant to be put into use. It has a strong product orientation, and in fact one of the most commonly asked questions from sponsors may be "What is the product of this research?" One product of applied research is its findings, which may be disseminated in presentations or publications. Sponsor-requested presentations may occur at several points during the course of a research program. In fact, some research contracts or grants are funded incrementally, and continuation is dependent on demonstrating satisfactory results at yearly program reviews.

Research presentations provide one way for the government sponsor to garner or maintain project support from various users or stakeholders, who may include not only higher-level program managers in their own organization but also higher-level program managers in the funding office or funding source and representatives from the actual end-user organizations. The applied researcher presenting his or her research results to this audience may quickly learn that there may be various "agendas" in the room, only one small part of which may involve the results of his or her research. Again, guidance and support from the government researcher/sponsor is useful in this setting.

Research projects may also result in publications in various outlets. Research contracts or grants typically call for a final technical report to be submitted to the funding organization. The sponsor may also request that this report be published as an in-house technical report. In practice, this means that the final report that is submitted at the end of a contract will go through several more revisions and approvals at various levels within the organization and eventually become published as an official technical report. The technical report will be disseminated through a distribution list (including the various stakeholders mentioned previously) and through the online repository, the Defense Technical Information Center's Scientific and Technical Information Network (STINET). A technical

report is a comprehensive write-up of research results but also should include guidelines or recommendations for applying the research results to the research problem or application. Remember, applied research starts with a problem and ends with recommendations for addressing that problem.

Research results may also be published in scientific journals, including those noted in the introduction of this chapter that encourage applied experimental submissions. However, note that publication or other dissemination of results outside of the government may require prior government approval. Any publication restrictions that may apply will be documented in the research contract or grant. This approval, if required, typically means a review of a draft manuscript by someone in the government organization prior to submission for publication. Approval is typically a formality, given that the research is unclassified rather than secret (applied experimental research in the social sciences is rarely classified as secret) and given that secrecy regarding the subject matter of the research is not deemed to be in the nation's interests. Therefore, although publication restrictions may be written into a contract or grant requiring prior government approval before publication, it would be extremely rare for publication to be disallowed. Finally, as we noted previously, applied research implies a partnership between the applied researcher and the project sponsor. In cases in which there may be an active and substantial contribution to a research project by the government program manager/researcher, it may be useful to establish an understanding for shared publication.

## VII SUMMARY

We have attempted to define applied experimental research and describe some of the common practical issues involved in conducting applied experimental research. We now consider if we were in Faraday's position and were approached by Gladstone or his modern equivalent with the question, "What is the value of applied experimental research?" Our answer, in keeping with the tone of this chapter, would address both the practical and the theoretical value of applied experimental research. At a practical level, applied experimental research can solve problems that people care about. Howell (1998) notes that "when properly planned, theoretically grounded, carefully managed, competently executed, and adequately funded, behavioral science is as capable of yielding solutions to significant real-world problems as are the physical or biological or any other sciences" (p. 424).

One example of a successful practical application of applied experimental research involved a commercial aviation accident that occurred on July 19, 1989. United Airlines Flight 232 experienced the failure of an engine and complete loss of hydraulic pressure, leaving the airplane with no flight controls. The airplane crashed during an attempted landing at Sioux City, Iowa, and 111 of the 296 passengers and crew members were fatally injured. In this case, the fact that the flight crew were able to bring the airplane down under some measure

of control was seen by most experts as just short of a miracle. Moreover, the [National Transportation Safety Board \(1990\)](#) noted in its investigation report that the flight crew of United 232 had recently received crew resource management (CRM) training, and this may have contributed to the outcome in which 185 lives were saved. Crew resource management training was developed from research applying principles drawn from small group research to the aircrew environment (see [Salas, Bowers, & Edens, 2001](#)). Thus, applied experimental research can result in practical applications that make a difference in people's lives.

At a theoretical level, applied experimental research serves as a bridge between theory and application. It serves as a major component of [Berger's \(1988\)](#) concept of a *theoretical research program*, consisting of theory, theoretical research to extend and elaborate theory, and applied research that grounds and further drives theory. It has been implicated in cyclic models that link theory, testing, and application in a recurring cycle or loop. As [Howell \(1998\)](#) noted, science is cumulative, and applied experimental research can lead to further refinements, further applications, and a deeper understanding of social phenomena.

## REFERENCES

- Ackerman, P. L. (2002). Editorial. *Journal of Experimental Psychology: Applied*, *8*, 3–5.
- Baum, A. (2005). Author guidelines. *Journal of Applied Social Psychology*. Available at, <http://www.blackwellpublishing.com/submit.asp?ref=0021-9029>.
- Beattie, G., & Shovelton, H. (1999). Mapping the range of information contained in the iconic hand gestures that accompany spontaneous speech. *Journal of Language & Social Psychology*, *18*, 438–462.
- Benjamin, L. T. (2006). Hugo Münsterberg's attack on the application of scientific psychology. *Journal of Applied Psychology*, *91*, 414–425.
- Berger, J. (1988). Directions in expectation states research. In M. Webster, Jr., & M. Foschi (Eds.), *Status generalization: New theory and research* (pp. 450–474). Stanford, CA: Stanford University Press.
- Bryan, W. L., & Harter, N. (1899). Studies on the telegraphic language: The acquisition of a hierarchy of habits. *Psychological Review*, *6*, 346–375.
- Bush, V. (1945). *Science: The endless frontier. A report to the President by Vannevar Bush, Director of the Office of Scientific Research and Development*. Washington, DC: U.S. Government Printing Office. Available at, <http://www.nsf.gov/od/lpa/nsf50/vbush1945.htm>.
- Cohen, B. P. (1989). *Developing sociological knowledge: Theory and method* (2nd ed.). Chicago: Nelson-Hall.
- Driskell, J. E., & Olmstead, B. (1989). Psychology and the military: Research applications and trends. *American Psychologist*, *44*, 43–54.
- Driskell, J. E., & Radtke, P. H. (2003). The effect of gesture on speech production and comprehension. *Human Factors*, *45*, 445–454.
- Driskell, J. E., & Salas, E. (1992). Can you study real teams in contrived settings? The value of small group research to understanding teams. In R. Swezey, & E. Salas (Eds.), *Teams: Their training and performance* (pp. 101–124). Norwood, NJ: Ablex.
- Freyd, M. (1926). What is applied psychology? *Psychological Bulletin*, *33*, 308–314.

- Gerstein, D. R., Luce, R. D., Smelser, N. J., & Sperlich, S. (Eds.). (1988). *The behavioral and social sciences: Achievements and opportunities*. Washington, DC: National Academy Press.
- Hackman, J. R., & Morris, C. G. (1975). Group tasks, group interaction process, and group performance effectiveness: A review and proposed integration. *Advances in Experimental Social Psychology*, 8, 45–99.
- Haverkamp, B. E. (2005). Ethical perspectives on qualitative research in applied psychology. *Journal of Counseling Psychology*, 52, 146–155.
- Howell, W. C. (1998). When applied research works. In J. A. Cannon-Bowers, & E. Salas (Eds.), *Making decisions under stress: Implications for individual and team training* (pp. 415–425). Washington, DC: American Psychological Association.
- Ilgen, D. R. (1986). Laboratory research: A question of when, not if. In E. W. Locke (Ed.), *Generalizing from laboratory to field settings* (pp. 257–267). Lexington, MA: Lexington Books.
- Kendon, A. (1983). Gesture and speech: How they interact. In J. M. Weimann, & R. P. Harrison (Eds.), *Nonverbal interaction* (pp. 13–45). Beverly Hills, CA: Sage.
- Krauss, R. M. (1998). Why do we gesture when we speak? *Current Directions in Psychological Science*, 7, 54–60.
- Krauss, R. M., Chen, Y., & Chawla, P. (1996). Nonverbal behavior and nonverbal communication: What do conversational hand gestures tell us? *Advances in Experimental Social Psychology*, 28, 389–450.
- Langewiesche, W. (2003, November). *Columbia's last flight*: The inside story of the investigation—and the catastrophe it laid bare. *Atlantic Monthly Online*. Available at, <http://www.theatlantic.com/issues/2003/11/langewiesche.htm>.
- Lewin, K. (1951). *Field theory in social science*. New York: Harper (Original work published 1944).
- Melton, A. W. (1952). Military requirements for the systematic study of psychological variables. In J. C. Flanagan (Ed.), *Psychology in the world emergency* (pp. 117–136). Pittsburgh, PA: University of Pittsburgh Press.
- Milanovich, D., Driskell, J. E., Stout, R. J., & Salas, E. (1998). Status and cockpit dynamics: A review and empirical study. *Group Dynamics*, 2, 155–167.
- Mook, D. G. (1983). In defense of external invalidity. *American Psychologist*, 38, 379–387.
- Münsterberg, H. (1898). The danger from experimental psychology. *Atlantic Monthly*, 81, 159–167.
- Münsterberg, H. (1914). *Psychology general and applied*. New York: Appleton.
- Murphy, K. R. (2002). Editorial. *Journal of Applied Psychology*, 87, 1019.
- National Science Foundation. (2013). *Federal funds for research and development: Fiscal years 2010–12 (NSF 13-326)*. Arlington, VA: National Center for Science and Engineering Statistics, National Science Foundation.
- National Transportation Safety Board. (1990). *Aircraft accident report: United Airlines DC-10-10 engine explosion and landing at Sioux City, Iowa (NTSB/AAR-90/06)*. Washington, DC: Author.
- Salas, E., Bowers, C. A., & Edens, E. (2001). *Improving teamwork in organizations: Applications of resource management training*. Mahwah, NJ: Erlbaum.
- Scott, W. D. (1903). *The theory of advertising*. Boston: Small & Maynard.
- Stouffer, S. A., Lumsdaine, A. A., Lumsdaine, M. H., Williams, R. M., Smith, M. B., Janis, I. L., et al. (1949). *The American soldier: Combat and its aftermath*. Princeton, NJ: Princeton University Press.
- Taylor, F. W. (1903). Group management. *Transactions of the American Society of Mechanical Engineers*, 24, 1337–1480.
- Thorndike, E. K. (1903). *Educational psychology*. New York: Lemcke & Buechmen.

- Triplett, N. (1898). The dynamogenic factors in pacemaking and competition. *American Journal of Psychology*, 9, 507–533.
- Turner, J. H. (1998). Must sociological theory and sociological practice be so far apart? A polemical answer. *Sociological Perspectives*, 41, 243–258.
- van Strien, P. J. (1998). Early applied psychology between essentialism and pragmatism: The dynamics of theory, tools, and clients. *History of Psychology*, 3, 205–234.
- Vaughan, D. (1996). *The Challenger launch decision: Risky technology, culture, and deviance at NASA*. Chicago: University of Chicago Press.
- Webster, M., & Kervin, J. B. (1971). Artificiality in experimental psychology. *Canadian Review of Sociology and Anthropology*, 8, 263–272.
- Webster, M., & Whitmeyer, J. (2001). Applications of theories of group processes. *Sociological Theory*, 19, 250–270.
- Zedeck, S. (2003). Editorial. *Journal of Applied Psychology*, 88, 3–5.

## Chapter 21

# Funding Experiments, Writing Proposals

Murray Webster, Jr.

*University of North Carolina—Charlotte, Charlotte, North Carolina*

### I WHY WRITE A PROPOSAL?

Times have changed. Until approximately the 1950s, most aspects of contemporary research in social science had not yet appeared. Many social scientists spent their lives teaching; they never even collected data. Some of the few who did research relied on government documents found in libraries when they needed information. Others reported on settings they observed themselves. Most often, publication meant producing interpretive theoretical discussions. Occasional small-scale investigations occurred relying on unpaid student assistance, even unpaid participation, in early experimental research. However, research in those days bore little resemblance to the social science research enterprise today.

Today, research is not optional; it is expected as part of the responsibilities of virtually all social scientists, whether employed in higher education, government, industry, nonprofit institutions, or the military. (Colleges and universities expect teaching and service as well as research, but in other settings, research is likely to be nearly 100% of the job description for a social scientist holding an advanced degree.) Furthermore, contemporary empirical research requires extensive planning, difficult and expensive data collection, a team of researchers and research assistants, sophisticated analytic tools, computers and other technology, and data archiving and retrieval. In short, research today differs both quantitatively and qualitatively from that in earlier days. Its character is very different and there is much more of it. The changes make research much more expensive.

Where does the money come from? There are actually many more sources of funding than most people realize. The largest source, not surprisingly, is the U.S. government. However, the federal government is far from the only source. Most national governments fund research, including governments of very poor countries that have little or no surplus money. Every branch of the U.S. government—the Departments of State and Agriculture, the FBI and CIA, Defense, Homeland

Security, Treasury, etc.—has programs to support research as part of its mission. The so-called “independent agencies,” including the National Science Foundation (NSF), the National Institutes of Health (NIH), and the National Institute of Justice (NIJ), exist explicitly to support research. In the United States, every government level below the federal—state, county, and city—also funds research. Private foundations large and small fund research. Most people have heard of the large Bill and Melinda Gates, Ford, and Carnegie Foundations, but they might be surprised at how many funding agencies exist. It is a rare community and an even more unusual topic that does not have a funding foundation dedicated to supporting it.

Those sources provide what is called “external funding”—that is, a source external to the institution that receives the funding, such as when the NSF awards a grant to an investigator at a university. Internal funding is another source of funding. Although few universities are rich enough to support much of their faculty research, most offer small grants, often referred to as “seed money,” in hopes that recipients will grow larger external grants from the initial results. For all of the other institutions mentioned, internal funding constitutes the majority of their expenses. Drug companies and other for-profit corporations fund their own research internally, as do nonprofit organizations. However, both for-profit and nonprofit research firms are heavily engaged in writing proposals to the funding agencies that provide most of their budgets. Federal government agencies such as NIH fund internal research along with funding much research done by others; county social service agencies may pay for studies by their own employees or contract employees.

Research funding is so widely available and so differentiated that we may say with only slight exaggeration that whatever topic you might wish to investigate, there is someone who wants to support you. The first task is to find out who that might be, and the second is to convince the funding agency that you are a better candidate to invest in than someone else might be. Finding the funders is accomplished with a little investigation online, at a library, or, most effectively, with the help of someone from a university’s or a corporation’s research office, sometimes called an Office of Sponsored Programs and Research. Research offices have large electronic listings of funding agents, the types of research they seek to support, amounts of support, how to request it, and other useful information. They include funding alerts from agencies and a variety of databases showing research support. A good research officer knows all the published sources, as well as having access to much more information received in the mail, found in professional journals and other publications, and gleaned at meetings of professional research administrator societies.

The second task is to develop a research proposal. A proposal is an offer to do a particular research project in return for monetary support. As any economist could tell you, if an institution gives away money, people are likely to ask for it. Your task is to make your proposal more appealing than anyone else’s.

Although it may be obvious that externally funded research requires proposals, internally funded projects usually require proposals also.

Research support falls into two broad categories, depending on how much discretion a researcher will have over its topic, methods, personnel, and outcomes. Sometimes an investigator thinks a topic is worth investigating and the proposal must convince his or her supervisors to spend the money on it; that is called an “investigator-initiated project.” Other times a corporation or a supervisor wants something investigated and asks teams within the agency and/or outsiders to design competitive proposals that will determine who gets to do the work. Investigator-initiated projects usually are supported by “grants”; the term means that the funding agency solicits proposals to do research, perhaps within a broad area (e.g., something related to employment opportunities of women), but otherwise does not specify details. The other kind of award is a “contract.” This term applies when a funding agency specifies the outcome (or “product”) expected and may also specify personnel, research methods, and other details of the work (e.g., to design ways for new female hires at XYZ corporation to progress as fast or faster than men in their careers).

Experimental research is supported by both grants and contracts, although the greater definition of contract research means that much of this chapter contains information on preparing grant proposals. Although much of this information should be helpful for writing proposals to do any kind of research, I focus special attention on the needs of experimental proposals.

Besides the obvious benefit (finding money to support research), proposal writing confers a number of indirect benefits. It helps focus and clarify ideas for oneself and others. Anyone who has tried to explain something—in a presentation to colleagues or to students in a class, for instance—knows that explaining requires and fosters deep understanding. Similarly, there is a major difference between having a good idea and developing that idea into a proposal so that others can appreciate its value. For that reason, ideas in proposals tend to be much better than ideas that live only in someone’s head.

Proposals receive feedback from other competent individuals. Although anyone can ask others to give feedback on an idea, or even to read a paper and comment on it, most of us find out quickly that that is asking quite a bit. Most of the time, most competent people are busy with their own work and they hesitate to take time to think deeply about someone else’s ideas. Reading a paper for someone is quite an imposition on one’s time and work habits, and thinking about it and offering suggestions for improvement multiplies the time and effort required. Yet proposal reviewers have agreed to do just that. Often, the review process generates the most thoughtful feedback available.

Along with supporting research and thereby advancing knowledge, successful proposals generate support for research assistants. Not only do research assistants benefit from a salary but also funded projects often provide an important way for them to learn techniques and strategies of knowledge production. For graduate students, research involvement is invaluable; for research

assistants in non-university settings, it expands their skill sets, eventually enhancing earning capacity and job mobility.

There are many other indirect benefits of proposal writing, including making a researcher more visible in the discipline, reflecting favorably on one's university or firm, permitting an experimenter to pursue his or her interests rather than tasks assigned by someone else, and even leading to advancement of knowledge. Proposal writing is destined to be a very important part of the professional activity of most social scientists, and certainly is so for every experimenter. Everyone needs to know how to write a proposal, and successful researchers need to know how to do it well. Yet proposals are competitive. Most of the federal funding agencies can support fewer than 25% of the proposals they receive. The situation is a little better with some private foundations though certainly not with all of them. I do not know of any agency that funds more than 50% of the proposals it receives, although there might be some that do.

Proposals will be read by reviewers who are other professionals from the discipline of the proposal writer. Sometimes, as with the larger federal agencies, reviewers come from the national pool of scholars. Other times, as with smaller agencies and many private foundations, reviewers are members of the funding agency—in particular, the program officers who receive proposals. Foundation proposals are often also read by members of the foundation's board of directors to ensure concordance with the foundation's purposes. Successful proposals must appeal to reviewers, in other words. Success is not magic. It is competitive, and it is surprisingly fair. Researchers who succeed at getting proposals funded write proposals that reviewers regard highly. That is really the sum of the matter.

Because proposal writing is competitive, it makes sense to invest some time developing winning strategies and techniques. Short of winning a lottery or marrying extremely fortunately, researchers really have no other choice than to enter the competition. As with other areas of life, success is never guaranteed and is not always achieved. However, preparation, effort, perseverance, learning from one's own and others' experiences, and viewing any setbacks as temporary can enormously improve the odds of success.

## **II SPECIAL PROBLEMS OF EXPERIMENTAL RESEARCH IN THE SOCIAL SCIENCES**

Experimental researchers may face particular hurdles in explaining this kind of research to funding agencies. The most common misunderstandings include concerns that experiments have low external validity or that results gained from experiments on college sophomores do not apply to individuals in other settings, but there are others. It is important for experimentalists to anticipate them where possible and try to keep them from harming a proposal's chances for success.

Misunderstandings and concerns about experiments often are misplaced, and they may come from a reader's ignorance about research methods and the relation of theory and research. However, a research proposal is not really the

place to try to educate one's colleagues (even assuming they want to be further educated). The best approach usually is to construct a scientifically valid proposal and emphasize what knowledge it can produce, being careful not to inflate claims. It is important to avoid appearing arrogant, as if an experimenter believes that this method produces better knowledge than other methods. A modest approach, including, when appropriate, acknowledgment that there are limits to the knowledge that the proposed experiments will produce, often defuses readers' concerns about supposed experimentalist arrogance.

An experimentalist can anticipate and sometimes avoid unfounded concerns about generality of findings by having a good grasp of the role experimental research plays in developing social science knowledge. Other chapters of this book should be useful here. Anyone who regularly conducts experimental research will have had numerous opportunities to discuss its philosophical foundations with other social scientists. The way an experimentalist writes a proposal can benefit from what he or she learns in those discussions.

Some misplaced criticisms appear regularly enough that it is wise for experimenters to anticipate them and, to the extent possible, reduce their harmful impact on one's proposal. Following are some common concerns and some suggestions that may be especially useful for experimental proposals:

- Experiments are artificial, in the sense that they are created by humans rather than by nature. This is an advantage of the method because it makes strong tests possible of theoretical predictions. However, that fact is not widely appreciated outside experimental circles. In a proposal, there is no need to apologize for artificiality, and it is useless to pretend that it is not a fact. Any attempt to claim that an experimental design seems "natural" or "realistic" misses both audiences. Experimentalists see naturalism as a design weakness; those hostile to experiments see any attempts to create naturalism as inevitably falling short of the mark. Explaining an experimental design in terms of the conditions the experimenter intends to create, and why those conditions will permit tests of the project's hypotheses, is a much better strategy than entering into artificiality debates in the proposal.
- Experiments should not be contrasted to "the real world." An experiment is as real as any other setting, and participants engage in experiments as fully as they do in any other interaction. A proposal to do this sort of research deserves a serious description of the design. The research should not be treated like a game or some other situation in which participants know their actions do not really count.
- Experiments are one research method—not the only method. They are well suited for certain kinds of questions, especially for testing hypotheses derived from well-developed theories. Other methods are better suited for other kinds of problems. An experimental proposal should describe why the experiment constitutes a good way to get knowledge in the particular case described, but there is little merit in triggering a discussion of what kind of research method is the best one in general.

- Experimental research is difficult to do, if done well. People who do not conduct experiments sometimes do not appreciate how much effort and planning go into a well-designed experimental study. They may not understand why experimental research takes as long as it does, or costs what it does, unless the proposal carefully describes all parts of the empirical work. To simply write about studying the effect of some independent variable trivializes the design. An experiment reveals the effects of an independent variable under certain initial conditions, and with certain operational measures, and many aspects of the design need pretesting to be sure they work as intended. The proposal should contain enough information that uninitiated readers can appreciate the many steps in the research. Every step in the research process incurs costs for personnel and supplies, so the proposal should show why those costs are justified. If incentives are necessary for subjects' payments, for instance, provide information about why they are important.
- Finally, a few social scientists think that experimenters are likely to mistreat human subjects. Perhaps this suspicion comes from the widespread showing of videotapes of Milgram's shock experiments or Zimbardo's prison research in introductory classes. Although suspecting the motives and ethics of experimenters is unfair to and misplaced for every experimenter I have known, it is wise to think about the suspicion when writing a proposal. This is especially important when the design involves any deception, as many social science experiments do. The author should mention steps taken for protection of human subjects, including institutional review board approval (discussed in Chapter 2). The design should go beyond the minimum required by law, however, and the proposal should describe full explanations to participants at the end of the work, offering to answer any questions a participant may have about the experiment, any other design features to reduce anxiety, the fact that participants will be volunteers and whether they receive pay or other remuneration, and whatever else shows that the investigator recognizes the responsibility to deal ethically with people.

Although these are some of the most common misunderstandings and prejudices about experimental work, there are many others. I have found that the best way to deal with all of them in a proposal is to take great care in explaining all aspects of the research design and show how each element is related to creating important conditions or gaining a significant piece of empirical knowledge. Of course, that is good advice for any research proposal. However, part of the special challenge of doing experiments is that not everyone has a good understanding of this kind of work, and not everyone automatically presumes experimenters are concerned with human welfare. Nevertheless, the job of a proposal writer is to make the value of the work and the motives of the researcher clear, even to readers who may not feel disposed toward this sort of work or who misunderstand crucial features of it. An investigator can do a better job of that if he or she thinks about what kinds of presumptions and background knowledge others in the discipline are likely to bring to the table when they read a proposal.

### III THE STRUCTURE OF RESEARCH FUNDING; ROLES AND ROLE BEHAVIORS AND SOME TERMS

For understanding the research enterprise, it is useful to begin by making explicit the larger picture of institutional structures and roles that I will allude to throughout this chapter. Researchers can be more successful in writing proposals and getting funding if they are clear about the structures and role behaviors of everyone involved in the process.

At the highest level, two kinds of institutional structures are important. One is the university or research organization at which the researcher works. The other is the governmental or private agency that pays for the research.

Although there are several differences between working at a research organization or a university, one thing applies to both institutions. A research grant or contract is a legal agreement between parties that are organizations, not individuals. In other words, if the NSF makes a grant after you submit a proposal from the University of North Carolina (UNC), that grant is part of a legal agreement between the U.S. government and UNC. This means, for instance, that if you change jobs and leave UNC, UNC has the option of deciding to send the grant funds with you to your new job or to retain them and appoint another competent person to fulfill the research. You may hope UNC will do the former, but it is not automatic and you should investigate if this circumstance might apply to you.

However, for all of us, the status of the agreement has a significant benefit, which is that the institution, not the investigator, must fulfill many legal requirements that most researchers are not equipped to fulfill. Among them, Congress requires that contractors guarantee a drug-free workplace. Although a researcher probably will supervise his or her research staff to be sure drugs are not part of the picture, no faculty member is able to supervise the entire campus to make sure it is drug-free. The nature of the legal agreement means that it is the university's problem, not yours, to certify.

There are other legal requirements with which employers must comply. Some of these include employment access to members of underrepresented groups and individuals with physical limitations, review by an institutional board for the protection of human participants or animal subjects, and compliance with various payroll laws such as minimum wages and medical benefits. Researchers do not need to keep up with all of these requirements, but they do need to understand that before they can submit a proposal from their employer, someone is going to review it to be sure nothing in it will violate a federal, state, or municipal law.

Institutions that provide money for research are called funding agencies. As noted previously, they include both government and private agencies. They provide money as part of their mission. Government agencies all have missions, as noted previously. Private agencies often want to affect society or some of society's institutions such as schools or families. Thus, funding decisions are made first and foremost in terms of whether a particular project seems to have a

reasonable chance of furthering the funding agency's mission, and that justification must appear in every proposal, either implicitly or explicitly. Researchers sometimes add irrelevant justifications, such as they or someone they propose to pay "deserves" the payment, or the researcher is tired and needs some relief from his or her normal work duties. Those things just waste space in a proposal because the agency is prevented (by law, in the case of government agencies; by internal regulation for private agencies) from taking them into account.

Both government and private funding agencies provide money in the two categories described in [Section I](#). Excellent experimental work has been supported both by grants and by contracts in recent years.

Within a funding agency, two roles are important: program officer and reviewer. The program officer is less well understood by many researchers, many of whom have reviewed manuscripts and even proposals, but program officers are very important and it is worthwhile understanding something about their roles and role relationships to researchers.

Most program officers are highly trained; the large majority of them have doctoral degrees in the field they oversee or in related fields. They read the scholarly journals and keep up with research; they are likely to know the work of many of the scholars in various fields. Program officers know what good research looks like, and they have an excellent understanding of the relations between theory and research design. Some program officers serve temporarily at funding agencies and then return to their universities or research organizations, whereas others make their careers within funding agencies. All, however, are as well qualified in their disciplines as are university faculty members.

The program officer is the "face" of a program, its public representative. Reviewers, by comparison, are less likely to be visible to outsiders, and they may wish to be anonymous to avoid inappropriate contact from researchers seeking privileged access to funding. (I discuss reviewers in more detail later.)

However, while a program officer represents a program to a researcher, the really important relationship flows the opposite way. The best program officers conceive their job as representing members of the discipline to the funding agency—that is, advocating for the needs and interests of the researchers. It is a common mistake to think of a program officer as a gatekeeper, keeping a researcher away from the money that he or she needs. Far more often, a program officer is more than sympathetic and looks for (and often finds) ways to help researchers achieve their goals.

I once heard a program officer summarize her relationship to researchers in this way:

*My job is to place 6 million dollars with researchers in my discipline. I will do that; no question. Your job is to make it easy for me to place some of that money with you.*

Of course, much of this chapter is devoted to offering information and suggestions to help experimental social scientists do that job.

A researcher who submits a proposal is an “applicant.” It is not important to remember that term, except to remember that “supplicant” would be inappropriate. In other words, the role relationship of program officer and applicant is an exchange in which both parties hope negotiations will succeed. Do not think of a program officer as hoarding money; the program officer is going to allocate whatever budget he or she has. Young applicants sometimes appear to regress and become adolescents asking a parent for an allowance. Whining, excessive politeness, and flattery—all techniques that may have worked within the family—are unprofessional and embarrassing for all. Avoid them.

Perhaps more commonly, applicants sometimes act as if all they have to do is convince a program officer to give them money. That is almost never true. For many reasons, program officers want to fund projects that others in the discipline—the reviewers—recommend funding. Although program officers make the final decisions, and for any of several good reasons those may not coincide perfectly with reviewers’ recommendations, program officers do not act alone. Nagging a program officer and other pressure tactics do not work because, in addition to making a researcher seem unpleasant to deal with, the program officer’s views are only part of the equation.

Previously, I wrote that the program officer represents researchers (you and me) to the funding agency. Part of our responsibility, thus, is to avoid making the program officer’s job more difficult than it needs to be. Social scientists, especially the young, inadvertently cause trouble for program officers in government agencies when the titles of their proposals raise flags at Congress. Although social science has some great friends in Congress, there also are a few who view us with suspicion, and those few often ask federal agencies to give them lists of proposals from our disciplines. They are looking for reasons to cut funding for the social sciences, and inflammatory titles—no matter how cute the researcher thinks they may be—are dangerous to everyone. It is in everyone’s interest for applicants to think about possible consequences of what we write in proposals that go to the government—in fact, in everything we write for all audiences.

In [Section I](#), I suggested that a researcher begin the quest with the research office at her or his institution. Research officers can help clarify topics and identify possible funding agencies. Once a researcher has settled on a few agencies, based on their mission statements and other information from the research office, she or he should contact those agencies. As noted previously, it would be a mistake to contact a funding agency without first learning its mission and other aspects of its functioning from the research office—at best, a waste of time, and at worst an embarrassing ignorance that can come back to haunt the researcher.

The program officer is the first point of contact with a funding agency for a prospective applicant. Phone calls and e-mails constitute the initial contacts in which a researcher can confirm understandings of the regulations of the agency, its mission, and other important details of proposal submission. This is a good time for a prospective applicant to outline (briefly) the topic of the contemplated proposal and hear if the program officer has any thoughts. With large agencies

such as the federal agencies, program officers often can suggest exactly the right program to target a proposal to, and they can clear up any number of potential misunderstandings right at the outset. It is well worth a researcher's time to initiate this contact.

Is the program officer too busy to bother with someone as insignificant as I am? Program officers certainly are busy. However, because of their roles, and by personal inclination, they really like to talk with us. Most program officers are passionate about research and their agency's mission. When you understand that a large part of a program officer's day, like almost everyone else's, is spent processing paper and attending meetings, you can appreciate how welcome an interlude to discuss research is for her or him. Yes, she is busy and yes it is an interruption when a researcher calls, but a good intellectual discussion (a brief one, remember) can restore a program officer's energy for the remainder of the day. Call.

There is one caveat here. Your call or e-mail to a program officer deserves to be well thought out, just as any other conference. It would be rude to call up anyone at work and free associate for a while. The sort of call a program officer dreads begins with "I would like to bounce some ideas off you." Program officers do not like to bounce.

Now that we have some of the basic facts about relations between universities and research organizations and funding agencies, about roles and role relationships within and among the institutions, we interpret how these play out in the life of experimental social scientists seeking funding for their work.

## IV RESEARCH PROGRAMS AND PROPOSALS

A research project, like a proposal, should be seen as an element in a larger plan or pattern. A proposal is not an end in itself; rather, it facilitates doing research and publication, and all the indirect benefits alluded to previously. A research career, ideally, will include many proposals at different times for different purposes. Some of them will succeed, leading to funding; others may not. Wherever possible, writing a multiyear plan for research topics, considering what to do first, what follows, what is next, and the like, makes for a more coherent career with less wasted time.

If you are the sort of person who likes planning, it may make sense to draw a time line for the next 5 years, for example. What research topics would you like to address in that time? What experiments will help do them? How long will each experiment take to complete? When does a proposal have to be finished in order to support an experiment? Questions such as these can help someone to organize more efficiently, and in most cases, they also provide reality checks on vague or unrealistic plans. For instance, if it will take 2 years to write a proposal and conduct one experiment, it is not likely someone is going to do more than two such experiments in the next 5 years.

Another use of a time line is to help thinking about programs of sequential experiments. Often, results of one experiment suggest new experiments, or new experiments can build on established findings. A time line can show the structure of such an experimental program, and in doing that, it facilitates thinking in terms of programmatic development of knowledge about related topics.

Proposals should appear throughout a career in research. Just as it would be unfortunate to abandon proposal writing prematurely, it is a serious mistake to put off beginning it. A researcher should begin thinking about proposals at the very outset of his or her career. One reason for this is to develop proposal and research habits, which will be surprisingly difficult to pick up later. Another reason is that proposal reviewers tend to feel positively toward investigators at the start of their careers. If someone waits several years—for instance, until achieving tenure at a university—before writing the first proposal, he or she handicaps himself or herself. You can bet that person's proposal will be considered in competition with a proposal from someone else who is just starting her career. Reviewers then will ask, "How do we know this older person really is going to do the work?" After all, the record for the past 6 years may show no evidence or promise of research. Or reviewers may just want to help a beginning investigator: "Let's go with the young person and help her get her career started right." That kind of age discrimination may well be unfair, but we are not going to change that now. Better to start early and thus let reviewers see you as the new hotshot.

If you are one of the people who come to experimental methods late, or if you are writing your first proposal after some years developing other aspects of your career, the previous suggestion should not discourage you. A better way to use the information is to acknowledge the concern some readers might have and to explain what you were doing up to this point. Often, some of that experience is relevant to the proposed work, and pointing that out can strengthen the proposal at the same time that it explains why it did not appear earlier.

By the way, discouragement can appear at any time. Later, I address discouragement following proposal rejection. However, discouragement ahead of time is extremely self-defeating because it prevents even entering a competition. There are many good excuses and bad advice to discourage proposal writing. Funding agencies suffer budget cuts; someone may think he works in an unpopular area; there are many other demands competing for scarce time; older colleagues may not appreciate how crucial proposal writing has become since they began their careers. None is a good enough reason to stay away. If you look hard enough, it is possible to find good reasons not to do anything at all! But except for the few who spend their lives in semidarkened fortified rooms, most of the rest of us take chances and engage with life.

Presuming someone has an idea for a worthwhile experimental study, how should she or he proceed with the proposal process? Generally, cast as wide a net as possible to enlist others' help. Talk the ideas over with colleagues as much as they are willing to do, to find weak points, refine good points, and find

ways to explain things clearly. Invite a colleague to become a collaborator, especially if, as is often the case, that person brings talents complementary to your own. For instance, you might need someone with mathematical modeling skills, or if you are a new investigator, you might ask a more experienced experimenter to be part of the team.

After enlisting help with the project, take advantage of practical help from people whose professional skills are in the area of proposal development and management. Visit your organization's research office and find out potential funding agencies for your proposal. When you actually begin writing, ask the research office to help with editing and preparing parts of the proposal, such as the budget, that may be mysterious. Research officers are usually eager to help, and when they are good at their jobs, they can be invaluable.

## V PREPARATION FOR WRITING A PROPOSAL

Once you have settled on one or a few potential funding sources, study their proposal requirements. Some issue RFPs (requests for proposals). Others have regulations on proposals, and those may include forms to complete as requested. Know whom you are writing for and how they want it done. Two things are crucial. First, learn and comply with their regulations. Those often include submission dates, page limits and formatting, and other material such as an abbreviated list of publications. (In a lecture, I once mentioned to a graduate class that one of the federal agencies specified margin widths (at least 1¼ inches all around), paper color (white), and font (Times New Roman or Arial) and wanted the proposal fastened with a single staple no more than 1½ inches from the upper left corner. A student objected that seemed very picky. I responded that if you are asking for a quarter of a million dollars, it makes sense to ask in the way they want.) Second, learn and write to the funding agency's purpose or mission. Every institution has a mission statement; ask for it if they have not already provided a copy. For instance, the NSF supports (mostly) basic research to advance scientific knowledge. The NIH supports basic and applied research to improve the health of Americans. Do not submit a proposal for an experimental study of health behavior to NSF; do not propose an experiment with no connection to health to NIH. Private foundations have missions and goals that may change from time to time; they seek proposals to advance their goals. Show in your writing that you understand the funding agency's mission and how your proposed research can help achieve it.

Having completed the preliminary work, a researcher can begin to think seriously about planning the proposal. Here, it definitely helps to list items because there are a lot of them and most are interdependent. Who will be responsible for writing the proposal or the different parts of it? What resources will the writer(s) need—for instance, access to publications or consultation with others? What kinds of resources and permissions have to be in place before submitting the proposal? For instance, experiments require approval from institutional

review boards (IRBs; discussed in Chapter 2), and someone probably has to agree to provide space and time of the researchers and other workers on the project. How long will every part of the writing tasks take? Working backwards from the agency's submission date, identify how long each part of the process will take, allow a little extra for unanticipated delays such as someone being on vacation or problems with the university computing system, and that determines the start date for the writing.

Proposal writing is hard work, and doing it well requires setting aside considerable time for the task. Successful proposals require planning and work, for a very good reason: a research proposal has at least as much intellectual content as a paper submitted to a scholarly journal. It contains a brief summary of relevant existing work, describes the new proposed research, shows evidence that it meets current standards of evidence, tells what it will mean in terms of changing the state of knowledge, and tells what comes next. Furthermore, a proposal is more demanding for several other reasons. Most important, it must stimulate enthusiasm. A reader must not only understand the quality of thought in it but also come to share the writer's belief that this work really ought to be done. A successful proposal conveys the writer's enthusiasm and brings some of it to the readers. Without enthusiasm, reviewers are not going to recommend putting an agency's scarce money into the work.

Previously, I noted that funding agencies provide guidance about what proposals to them should contain. Every proposal, however, has certain broad sections to it. Each section leads to the one following it. More important, the justification for each part of a proposal depends on what has been established in the preceding topic. A well-crafted proposal "flows" from an orienting statement through the various parts of doing the work to a conclusion in a way that seems, if not inevitable, at least reasonable and thoughtful.

## VI SECTIONS OF PROPOSALS

Proposals contain five large groupings of topics, or five parts. The five parts can be seen as analytic components. The actual sections of a proposal may be specified in the agency's RFP or its guidelines. The "parts," as that term is used here, are things that should be included in every proposal, whether or not they get labeled as sections of the proposal. Be sure the proposal attends to the issues of each part.

The first part is the proposal's overall *topic*, and it is probably the most important part of a proposal. A proposal's topic should appear early and clearly enough so that readers have a context for all that follows. For example, a proposal's topic might be to extend theories of exchange in networks—that is, to develop theoretical understanding of those processes. The topic statements of a proposal should help answer (or avoid) questions such as "Why are they telling me this?" at later stages of the proposal. That is, every sentence in a proposal should obviously help to deal with the topic as initially stated.

Second, every proposal contains one or more *research questions*. These deal with a part of the general topic just described. A research question is more specific than a general topic because no research could fully explain every question in a topic. For example, the proposal might contain a research question such as “Do participants who engage in negotiated exchange come to trust their partners as much as those engaging in reciprocal exchange?” (In this discussion, do not worry about what those terms mean; trust that the proposal defines them somewhere.) Research questions may be stated as hypotheses from explicit theory or other sources, or they may be stated with less background. Their crucial property is that they must be answerable from research. That point is not always appreciated; later, I mention some instances of unanswerable research questions. At the same time, a proposal must justify its research questions as significant, questions whose answers will help explain the area of the topic.

Third, every proposal contains a *research plan*. For an experimenter, this includes the experimental design and operations, as described in Chapter 1. The research plan should be tied to the research question, so a reader can clearly understand why doing this would produce answers to the questions. Contrasting conditions of the experiment should have a reasonable chance of showing whether the research question is to be answered “yes” or “no” or whether a research hypothesis is confirmed or disconfirmed. For example, Condition 1 will have people participate in negotiated exchanges, condition 2 will have them participate in reciprocal exchanges, and afterwards participants in both conditions will complete some measure of the degree of trust they feel toward their partners. The research plan is the heart of the proposal and receives the most space.

Fourth, a proposal identifies *personnel* who will conduct the research. This includes the principal investigator (PI) and any co-PIs, research assistants, consultants, and others. Of course, every person’s contributions to the project must be spelled out and their inclusion briefly justified. The ultimate justification for every person identified must be that she or he contributes to the research plan described previously.

Fifth, a proposal contains the related topics of *timetable* and *costs*. Some agencies want time lines; others accept descriptive text with tasks and estimated times for completion. Time is money because most of the cost of research is in hiring personnel. Proposal budgets often follow strict format restrictions. Most researchers are not familiar enough with all the restrictions to complete their own budgets, turning that task over to the skills of professionals in the research office.

One way to think about the five parts of a proposal is to recognize that readers will be evaluating each part. Does the topic deserve further research? That is, is there a foundation to the topic so that others will work on it, or has its value played out? Is the research question a good one? That is, has the writer found a question whose answer will advance understanding of the topic? Readers ask questions such as these of each part of a proposal. When a proposal shows that the PI has thought through the reasons at every stage, it tends to generate positive

impressions and even enthusiasm among readers. Such a proposal is likely to attain the enviable state of being considered “competitive.” Now let us expand each of the proposal’s parts and examine them in greater detail.

## A Topic

The topic of the proposal is the general theoretical context into which the research falls. Perhaps a PI plans to investigate power use in certain network structures or identity processes in mixed-gender groups. Whatever the area, the research must be placed into a research tradition. This is the place to show that the PI understands how others have thought about the topic and what related research has shown. Then the PI will go beyond what is known to propose new questions and answers. The background material is sometimes called a “focused literature review.” The purpose here is not, as often appears in journal articles, to display the encyclopedic knowledge of the writer. Rather, the purpose of reviewing literature is to tell a story of increasing understanding by describing work that leads up to what the PI will propose doing.

The topic of research affects everything that follows in a proposal. For experimenters, that topic is often something in group processes, broadly conceived. Whatever the topic is—network structures, status processes, organizational legitimacy, power relations in informal groups, identity processes, or anything else—the proposal ought to identify the topic early. Early statement sets the stage for the work to be described, setting context and suggesting possible specific questions. If readers understand the topic and know it is important, the research questions have a better chance of looking worthwhile also, and so on.

The topic also suggests the skill sets needed by qualified reviewers—namely people who have demonstrated their interest in and knowledge of this topic. Clearly stating the research topic increases the chances of getting qualified reviewers. If a program officer cannot determine what a proposal’s topic is, besides considering the proposal weak on that ground alone, it becomes difficult to know who would be an appropriate reviewer. It is in the PIs interest to ensure that the program officer can clearly determine who understands the proposal.

Do not overlook research topics that might have interest for more than one discipline. Interdisciplinary topics often are very appealing to reviewers because they can contribute to knowledge growth in multiple fields and may help create links among them. Writers of an interdisciplinary proposal often benefit from getting chances to appeal to, and receive funding recommendations from, more than one set of reviewers. At the same time, good interdisciplinary work is difficult because it requires good knowledge of a discipline beyond one’s own. Furthermore, the interdisciplinary character must be real. A proposal that simply says “both psychologists and political scientists will be interested in this research” can look naive or worse if it does not adequately review what has already been done in one of the disciplines.

As this point suggests, it would be foolish to choose a research topic for which the PI has little or no training or competence. Just because a topic is in the news or it looks important on its face does not make it a good choice for a proposal. For instance, the publicity given secondhand smoking, responsiveness to sexual predators, appeal of murderous ideologies, and dozens of others topics does not by itself make them good topics for everyone. Unless someone has demonstrated competence related to answering research questions related to the topic, the proposal is not going to fare well with reviewers or funding agencies.

## B Research Question

This is what the proposed research will, we hope, answer. It is a piece of a jigsaw puzzle whose overall picture is the research topic. In fact, the research question in a proposal is an important part of the topic, and the proposal must show why that is so. One of the questions reviewers will ask is whether the proposed research question will contribute meaningfully to the topic of the research.

When it is possible, a good way to present previous work is to use the image of a funnel. Begin with broad ideas appearing in general theories or even in topics of the theoretical social scientific classics. Then trace an idea through increasingly specific theoretical views and into empirical findings. Show what has been investigated and what has not yet. Then you can situate your proposed work in that body of theory and research. A variant idea is to trace two literatures through funneling and to show how your work can illuminate two traditions. Not every research tradition is orderly enough to permit funneled reviews, but where possible, it provides a strong heritage for the proposed work. Of course, any organization of a literature review depends on the reviewer's intellectual work in discovering and imposing order on a messy historical record.

For much experimental research, there is a more or less well-developed explicit theoretical foundation. If the research topic includes presentation of general principles, then the research question section may identify one or more of the principles that can be modified to deal with the subject of this proposal. Alternatively, the research questions may flow from a new model that is presented in this section. When the research questions come from explicit theoretical foundations, it is important to outline the reasons why these particular questions will be informative for assessing the theory presented.

There are several pitfalls in choosing research questions. First, of course, the questions must be answerable by empirical (or sometimes, by theoretical) research. Although that might seem obvious, it is surprising how often inexperienced researchers do not consider it. I once knew a student who wanted to answer the question, "Why is there war?" in her M.A. thesis. It is a wonderful question, but despite centuries of thought, nobody has come up with a convincing answer. Had that student actually begun that project, she would die of old age before coming to a scientifically valid answer.

Questions that might look more focused can also be unanswerable. For instance, questions such as “How should an organization set up a gender equality program?” do not have one empirical answer because organizations are complex and any program is likely to have both intended and unintended consequences. In other words, any findings would be conditional on particular, probably unmeasured, characteristics of the setting in which they were found. The “should” suggests moral questions that are beyond social science research. A more focused question such as “How much effect did a particular state law have on promotion rates of women and men in large and small corporations?” might be answerable with empirical research.

Another common mistake is to propose research for which an investigator already believes he or she knows the answer. Examples (nonexperimental) I have seen recently include the following: “Did the Administration create a ‘moral panic’ to justify the invasion of Iraq?” “Are Mexican immigrant women in New York hard-working and resourceful?” “Is community support helpful as released convicts adjust to living outside the institution?” and “Would it be a good thing if more communities developed recycling programs?” If you already know the answer, it is difficult to justify spending money to find the result. In other words, if an investigator is already convinced of an answer, is that the right person to design research that has a fair chance of finding disconfirmation? Most reviewers would conclude that the answer is no. Someone might wish to spend money to convince others of what the proposer believes, but such a proposal would go to a different kind of agency. Propaganda and dissemination are different activities than research.

A third kind of problematic question choice comes from the tenets of a philosophical position that had some currency in the social sciences recently: social constructionism. Sometimes the term denotes the study of ways in which meanings of interaction situations and social identities come to be shared by actors and ways some may try to control the meanings others assign. However, some extreme versions of social constructionism hold that what we call scientific knowledge is no more or less than what we can get scientists to agree to, and that any evidence is shaped by investigators’ theories. If someone truly believes that, there is no reason to do empirical research because the outcomes will do nothing more than confirm the investigator’s preexisting ideas. Again, if there is no possibility of disconfirming or modifying the state of knowledge by research, funding the research is not a good use of scarce dollars.

Research questions always need justification. Again, there are good and bad justifications. Here are some very bad justifications:

- “Nobody has yet studied ....” There are plenty of trivial topics that remain unstudied for the best of reasons: nobody cares to take the time.
- “This is related to my life (ethnic background, family, life experience).” Although this justification is seldom stated this baldly, readers can spot it

easily. However it may be disguised, this justification appears surprisingly often in social science and it is always a mistake. Reviewers' eyes glaze over.<sup>1</sup>

- “I found a hitherto overlooked setting (or group, or practice).” This one combines both of the preceding justifications. Does it persuade you? Probably not, and it certainly will not do well with reviewers, either.
- “This is really interesting to me.” Although understandable, this is an incomplete justification for doing the research. The real question is whether a topic is important to other researchers. Things I have recently discovered are often interesting to me, but often that fact just reflects my prior ignorance.

What makes a good justification? The best, probably the only, good justification is that the social science community is interested in a topic. In other words, published research has dealt with that topic or closely related topics, and what you propose to study—your research question—looks like it provides a missing piece to others' understanding of the topic. A research question is justified by being placed within the intellectual matrix of its discipline. The relevant community has seen this question and wondered about it. In other words, reviewers of your proposal can see that you are dealing with a topic they are interested in. They are on board intellectually. They are primed to want to know what your research will find, and they may even feel that hoped-for enthusiasm.

Picking research questions in this way requires viewing oneself as part of a community of scholars, people who have some broad interests in common and whose work sometimes helps the community at large understand things better. This view is very different from the hubris that leads some to think they are entirely unique thinkers, dealing with questions and answers nobody else has ever thought of. Science is most often a cumulative, shared enterprise. The lone wolves are far more likely to be crackpots than overlooked geniuses. If that view is dismaying, at least wait until you have proven you can contribute to understanding things other social scientists want to know, and then go on to prove your unique brilliance with unique research questions. (But remember that Isaac Newton—one of the greatest intellects ever, justly praised for his theoretical and experimental studies of gravity—thought he had just such unique theological understandings in his old age.)

How can you be sure that your reviewers have already dealt with your topic? That is the way reviewers are selected, in part. Funding agencies ask people whose own work has something in common with the proposed work to judge its merits. With any luck, one or more of your reviewers will note approvingly that you have reviewed their published research and are proposing to take the next

---

1. On the other hand, if your topic is *theoretically justified* (as it must be to have a chance of success), and if some aspect of your life makes you especially well-suited to investigate it, then by all means highlight that fact. For instance, if your experimental design requires some conditions of participants who speak Spanish as well as English, and if you are bilingual, be sure to point that fact out. That advantage makes you better qualified than a monolingual person to conduct this research.

step, intellectually speaking, to build on it. Work is significant precisely because it contributes to cumulative growth of knowledge, and the better a PI can demonstrate how work will do that, the more likely that work is to receive support.

Another useful question to keep in mind is what some have called “the so what factor.” In describing a research question, it is helpful occasionally to imagine someone asking, “So what?” Although most readers, like most listeners, are too polite to formulate that question quite so directly, it is always likely to be in the back of their minds. A writer has the responsibility of making the case for a project, and if few readers or listeners are convinced, that is probably due to a writer’s not doing a good job of it. Keep “So what?” in mind, and try to show in the writing that there are very good answers to that question.

## C Research Plan

This is the heart of a proposal—what the PI will do if awarded the money. The most important consideration here is whether the research plan has a good chance of answering the research questions identified previously. If you follow this research plan, will you get reasonably clear answers to the research questions?

The research plan is probably the section of a proposal that most social scientists are best prepared to write. Their training is precisely in how to design and conduct research. However, there are sometimes gaps between what a scientist knows and what that person writes in a research plan. The guiding directive for this part of a proposal is to explain everything; *never* assume “a reader will know what I mean.” Even if a reader does know how something ought to be done, the proposal has to show that the PI also knows how to do it. The amount of detail should be sufficient to allow a competent reader to assess the quality of design and operations.

In Chapter 1, we distinguished research design such as experimental conditions, independent and dependent variables, and characteristics of participants from research operations telling how design elements will actually be realized in an empirical setting. Design and operations must both appear in the research plan of a proposal.

In addition to describing independent and dependent variables, an experimental research plan makes clear how the proposed situation meets the scope conditions of a theory. The proposal describes experimental conditions and clarifies the ways they differ from each other. The operations tell who will be participants, whether they will be volunteers, whether they will be paid, and how they will be recruited. No federal agency will release funds until an IRB approves the design, and most private foundations impose similar requirements. It is important to show awareness of the importance of that review by including a sentence similar to the following: “This project has been submitted for review by our IRB, and if this proposal receives funding, no data will be collected until they have approved the design.”

Wherever possible, offer brief justifications for details. For instance, if an experimental proposal includes four conditions with  $N=25$  in each, that number should be justified by a power analysis showing that it is sufficient to find predicted differences. It is quite helpful to sketch what the data will look like and how you will analyze them. It is important to show that the PI has thought about such questions and has reasons for whatever decisions she or he has made in regard to them.

In experimental research, it is often possible to adapt a previously used basic design for the new research questions. When possible, that approach is highly desirable for scientific reasons described in other chapters. It also is desirable for a proposal because any previous uses of the design constitute pretests for the proposed modifications. Reviewers will know that most aspects of the design have already been used and they work successfully without unanticipated problems appearing. Whenever possible, building on existing research methods strengthens a proposal.

Experimental research usually produces patterns of results. Beyond stating predictions, a proposal needs to review how the predictions were derived and what results the PI will consider confirmation and disconfirmation. It is very helpful in a proposal to include discussions of what both predicted and unpredicted outcomes of an experiment would mean. For instance, a particular experiment might have predicted outcomes of conditions ordered  $1 > 2 > 3 > 4$  for the specified dependent variable. The proposal should make clear that is what is predicted and why that ordering is expected from the theoretical foundations. However, the proposal should go beyond explaining its predicted outcome; it also should mention what the data will look like if the new ideas are wrong and what other outcomes might mean. For instance, disconfirmation might produce  $1 = 2 = 3 = 4$ . The proposal should include a discussion of how disconfirmation might occur. It might mean, for instance, that the new theoretical ideas simply are wrong. It also might mean something is wrong with the empirical design or operations. A strong proposal will discuss possible meanings of disconfirmation and outline follow-up studies in case disconfirmation appears.

Many proposals neglect other possibilities, and often that is a mistake. Stronger proposals note some additional reasonable outcomes (e.g.,  $1 = 2 > 3 = 4$  or  $1 < 2 < 3 < 4$ ) and tell how those might be interpreted. Although not all possible outcomes mean something (other than poor design or operational error), some patterns do have meaning. Strong proposals show that the PI is aware of possible outcomes other than those predicted and those showing the predictions are clearly wrong, and the PI is prepared to follow up if they appear.

Finally, the research plan concludes with what will be known as the result of the work and what some next steps might be. This means returning from the operational level (the actual data) to the design and theoretical levels (what they mean and how they affect the state of knowledge about a topic). Next steps usually are further theoretical questions that may be pursued once results of the proposed experiments have come in and been evaluated.

## D Personnel

A proposal describes who will be associated with the project and in what capacity, usually in a separate section. Personnel includes the PI and any co-PIs, of course. In addition, there will probably be research assistants who help run experiments and interview participants and, possibly, consultants. The proposal should briefly indicate why individuals other than PIs have been chosen and what they will contribute to the project. Reviewers assess both roles (e.g., Why does this work contribute to the project?) and individuals (Is this the right person to do this work?). A strong proposal makes both questions easy to answer.

Reviewers assess whether the PIs' training or experience qualifies them to do the proposed research with a reasonable probability of doing it well. They also assess whether research assistants are well qualified and whether consultants are needed. As with other parts of the proposal, personnel should be justified. The fact that reviewers will assess the qualifications of the PI and others to do the work, which they sometimes call their track records, helps to explain why picking a research topic based on popular consciousness usually is not a good idea. Just because a topic is in the news, or because a PI thinks it is important, does not make that PI the right person to investigate it. If someone has no relevant training or experience studying, for example, global warming or sexual predators, reviewers are not likely to recommend funding a proposal on those topics by that person. The odds are that a newcomer does not know what others have already done and will overlook some well-known research pitfalls; and thus the chances of such a person making a useful contribution appear small.

## E Budget and Time Line

Previously, I suggested getting help with budget preparation from a professional in the institution's office of research. Budgets are complicated, and most researchers do not understand some parts of them. For instance, most university faculty do not know their university's indirect cost recovery rates and fringe benefits, or how those may vary between the academic year and the summer. Nor do they need to know those things. That is what research professionals know, along with other things. The main things a researcher does need to think about are all the tasks that need to be done to complete a project and how long each of them will take (called person hours, or person months, and, of course, directly reflected in costs of salaries and wages).

In the meeting with the research officer, the PI can describe who will do the work and how long it will take, and for research assistants and consultants, what the PI would like to pay them. The research officer knows the institution's policies regarding such esoteric topics as indirect cost recovery (sometimes called overhead), fringe benefits, tuition remissions, and other costs that a PI probably does not understand well. A research officer also can give the PI some reality checks about how much money a funding agency typically awards—for those

times when a PI thinks it will take a few million dollars to really do the job right or when he or she wants the funding agency to pay salary for 2 years so full time can be devoted to the research. (It does not happen.)

In thinking about a time line for doing the work, the starting point is a careful estimate of how long each part actually will take. To that, a PI should judiciously add some percentage because if anything unexpected occurs—and the unexpected almost always appears in research—the one thing we know is it will take time to adjust. Time, in most cases, is money. That is, research assistants and others will have to be paid during the adjustment period as well as doing the actual work. Remember also to include time for pretesting of experimental designs and for data analysis.

Now that we are considering money, I suggest looking back over the entire proposal as what it really is: a request for money. Read it again and focus completely on the money. This is not to be crass. Everyone knows PIs are motivated primarily by theoretical questions and the search for truth. But think of it from the funding agency's point of view. *The only thing they can do for the research is to write a check.* They cannot provide the statistical consultant a project might need or arrange laboratory space to conduct the experiments. All they can do is send money. This means that the proposal must convey that the PI is ready to go as soon as he or she receives the money. Anything not already in place other than money weakens a proposal, sometimes fatally. A funding agency official might wish to help with design and operations, but he or she cannot do that. All the official can do is approve sending money, so the proposal must convey that lack of money is the only thing holding back the important work it describes.

## VII SOME TIPS ON PROPOSAL PREPARATION AND WRITING STYLES

Proposal writing, like research, is an uncertain activity. Proposals and PIs are in competition with other proposals and PIs, and it is wise to presume that much of the competition will be very good indeed. Not every excellent proposal receives support, and there is no magic formula for writing a successful proposal. Still, here are some suggestions from someone who has seen a lot of proposals; I hope some of them are helpful.

Many proposals contain so much justification, especially of the topic, that they begin to look defensive. Do not waste space with excessive justifications. In fact, begin the proposal with what you want to do, and let justifications take care of themselves as you describe the intellectual background of the topic. A beginning sentence such as the following can focus a reader's attention on everything to follow: "I propose to conduct a six-condition experimental study to assess a new model of power at a distance in network exchanges." That is clear, and you can explain any possibly unknown terms such as "power at a distance" later in the proposal. At least the readers will know what you are proposing to do.

What I called a defensive beginning that unnecessarily handicaps a proposal is some version of “This general area is really, really important in people’s lives.” Many proposals begin with a paragraph or even a page or two that seem to have no purpose other than convincing a reader that the topic matters. That always looks desperate. Most of the time, reviewers are willing to grant that a topic is worthwhile or at least to suspend judgment until they have read your justification for the research questions (which you now know has little to do with saying you think it is important). Seeming desperate to be taken seriously is not a strong beginning for a proposal.

Develop ideas clearly and logically. Put the essence of the work at the beginning, and fill in gaps later. Reviewers want to know *what* you are going to do, and once you tell them that, they will be interested to know *why*. Doing things the other way around (why first; then what) invites confusion. Worse, it may invite irrelevant (from your point of view) ideas about other research questions as reviewers fantasize how they might study the topic.

Organize the writing to permit skimming. Include headings so readers can find things quickly if they want to review. When a reviewer gets to the end of a proposal, perhaps while reading the PI’s views of what the research means and what might be done next, the reviewer might wish to review exactly what are the different experimental conditions. Clear organization and boldface headings make them easier to find. Reviewers meet to discuss proposals at Study Sections of NIH and Advisory Panels of NSF, and also for some private foundations. During discussion, if someone asks a reviewer for a fact about your proposed research—why you propose to include only female participants in certain conditions, for instance—the reviewer must be able to find the answer quickly. Make it easy.

Get someone else to read and comment on your proposal before you submit it. This will help you find parts that are left out, justifications that need to be added, and details that are obscure. The best reader is someone whose research is in a field different than yours because that person is unlikely to mentally fill in the gaps when you have omitted something. Your schedule for proposal preparation has to include time for getting these opinions and revising the draft once you get them.

Aim for the style called “technical writing” in English departments. This style is far removed from the evocative style of creative writing or atmospheric essays. Technical writing is literal. To someone trained in creative writing, technical writing may be boring; that really is irrelevant. The only important criterion is clarity; a reader should have no doubt at all what you mean to say. Convey excitement through the quality of ideas rather than through the language used to express them.

One idea per sentence; one thought per paragraph. Each idea gets its own sentence, and no sentence tries to convey two ideas (as this one just did).

Avoid synonyms. Use one word for each idea, and use that word every time you intend that idea. Isn’t that boring? Maybe. Remember clarity is the criterion,

not “fun to read.” A proposal that sometimes says “stratification” and sometimes “social inequality” invites confusion. At the very least, a reader will wonder whether the two terms refer to the same thing and then look back and try to figure it out. After doing that a few times, this reader will be annoyed at the writer for making him or her put in that unnecessary effort.

Write simple declarative sentences. Readers should not have to wade through subordinate clauses and qualifications to get your meaning. Use active voice. Passives make it difficult to determine what you propose to do. Consider the difference between “It is hoped that respondents will see ...” and “I will instruct respondents that ....” The latter is a proposal; the former is a vague wish. Identify all acronyms and abbreviations the first time they appear. Avoid jargon as much as possible except where a term is so well-known and widely used that every competent social scientist would understand it in exactly the same way.

Proofread and ask someone else to proofread. *Do not* just use a spell checker. Spell checkers, helpful as they are, cannot identify misplaced words, misused words, or syntactical errors. They often cannot determine whether “its,” “it’s,” or “it’s” should be used; they do not know if you used “affect” where “effect” would have been correct. Any errors in syntax or grammar make a proposal look sloppy—definitely not the impression you wish to convey to readers.

Number the pages so readers can refer to them easily. Do not try to evade page restrictions by using tiny fonts or tiny margins. That really irritates reviewers, who often have many other proposals to read! Why would you want to irritate them when they are deciding whether to recommend your proposal for funding? Except in rare circumstances (e.g., when a funding agency requests them), avoid appendices of any sort. They may look like an attempt to evade page restrictions, and even if they do not, you cannot be sure all reviewers will read them. Make the proposal self-contained. Finally, follow all the funding agency’s regulations regarding treatment of human and animal subjects, data archiving, etc.

## VIII SOME STYLISTIC SUGGESTIONS

Using a technical writing style restricts some of the imprecision and distractions that can plague proposals. Here are some additional thoughts on how a proposal communicates ideas and expresses its writer. Consider this section as suggestions on avoiding problems that can make a writer appear differently from how he or she might wish to appear.

### A Title

The purpose of a proposal title is to tell what the proposal is about. Simple, right? Yet some writers treat a title as more of a teaser than a description. Advertisers and local news anchors may use teasers, but nobody really enjoys hearing them. Reviewers are going to read the proposal, no matter what its title.

There is nothing to be gained, and some potential losses, from writing a title that does not really tell the subject of the proposal. If the title accurately conveys what a proposal is about, readers' thoughts are primed to think about it. Readers are unlikely to imagine some other topic and have to return to your topic later, something that at least some will find irritating.

Closely related to how a title reads is the unfortunate tendency to believe that every title must contain a colon. Colon writers often try to be clever with one half of the title, while the other part tells what the project is really about. For examples, look at titles in most social science journals. On one side of the colon, usually on the left side, is a phrase the writer thought was cute. "Cute" has no place in a research proposal if the writer wishes to be taken for a serious scholar. If one part of a title tells what the proposal is about while the other part attempts to be cute, discard the cute part and use only the descriptive part. You never know when a cute part will antagonize a reader unnecessarily, and as I said, it gives an impression of the writer that does not help get funding. Eschew titular colonicity, or as a friend puts it, perform regular colonectomies.

## **B Avoid Distracting Points**

When we are emotionally involved with a topic, it is easy to throw in phrases and even whole paragraphs that tell how we feel but are irrelevant to proposed research. Try hard to stick to the point, and when you ask colleagues to read your proposal, ask them if they spot irrelevancies. For instance, I read a proposal to study gender bias in American Sign Language (ASL), which was overall a worthwhile topic. However, the author of that proposal could not resist dropping in sentences such as "However, through the years, deaf people have become an informed and empowered community." I am sure that is true; it just has nothing to do with how ASL or people using ASL handle gender.

Besides wasting space, irrelevancies always carry the potential danger of antagonizing a reader. Sociologists may drop in a phrase telling that they disapprove of some government or corporate policy or that has nothing to do with the research, and it is foolish to do so. Ask yourself how your comments will look to a reader whose political or other philosophies differ from your own. The answer should be obvious.

## **C Use Bullets and Summarize**

Often it is useful to enumerate a number of steps in research, or a number of variables or predictions, etc. Bulleted lists provide an economical way to do that. They also tell readers that they do not have to remember every item (e.g., the wording of every prediction) as long as they understand the category to which it belongs. It is quite likely a reader will want to return to your bulleted list to find specifics about a category, and this makes them easy to find.

Summaries are one of the main places to use bullets. At the end of a description of experimental conditions, bullet points might summarize the conditions so a reader carries them into a later section on, for instance, data interpretation. The end of a proposal is a good place to use bullets listing what the project will accomplish. Frequent summaries, whether or not they use bullets, are very helpful to readers.

## D Please Do Not ...

Following are some writing habits that can be off-putting in a proposal and probably in other writing as well. In a proposal, as frequently noted previously, bad habits can cost.

Do not write “see” if a reader cannot easily find the object it refers to. A particularly egregious instance occurs when a writer refers to his unpublished dissertation for important details, such as an experimental manipulation or measurement technique. In fact, it is probably wise to avoid the command to “see” anything. Even if a reference is to an article published in the largest journal in the discipline, readers simply do not have time to search out additional sources when they are reviewing. Every proposal must stand on its own.

On references, be sure everything cited in the proposal actually appears in the citation list. You might be surprised at how often the section called “References Cited” omits one or more of the papers a proposal text cites.

Although I have sometimes begun book chapters with a series of questions that I implicitly promise to answer in the chapter, questions usually are not a good way to begin a proposal. Questions are wordy, and if there is one fact true of all proposals, it is that their page limits always constrain PIs. Get right into the answers and let someone else imagine how to ask the questions that you are answering.

To make your proposal stand out and make the writing vivid, try hard to avoid clichés, of both expression and content. There is no point in writing the sentence “More research is needed” unless you can imagine sometimes writing its opposite. I recently sat on a review panel in which someone offered to buy lunch for any reviewer who had a proposal to do interviews or conduct focus groups *without* claiming they would be “in-depth.” He did not have to buy. Calling a book, a research project, or a publisher as “major” adds nothing. (When was the last time someone asked you for “a minor credit card”?) Finally, watch out for unnecessary sentences such as the following, from a proposal to study hospice usage, which are vacuous, redundant, and not even true: “Dying cuts across all categories of humanity. It is the one activity we all participate in, regardless of our gender, race, SES, or ethnicity.” (How about using language or socializing the young?)

Watch for and remove weasel words—those that negate part of a sentence. Common examples are “may,” “might,” and “necessarily.” To write that something “may” happen is to write that it might not happen. If both outcomes would

make the sentence true, it just wastes space in the proposal. Even worse, putting “may” in a hypothesis or prediction means no pattern of data could disconfirm it. To say that something is not “necessarily” true usually does not matter. What is important is whether it is true or not; the “necessity” part is not empirical.

If weasel words do not say anything, wildly inflated claims say too much. The problem with inflated claims is that readers can spot them, and rather than being impressed as the writer probably hopes, they do not take that part of the proposal seriously. A proposal that claims results will apply to all individuals of all ages, genders, psychological makeup, and ethnic backgrounds reveals philosophical and methodological naiveté. A proposal to employ both experimental and survey methods, and that claims the results of mixed methods will be so powerful that all subsequent research will have to use them, invites ridicule. (I have seen both claims.) Mixed methods and studying a range of diverse individuals are both good ideas. We all would like to think that our research will help others’ understanding of the world and of how to do research. However, to believe—and even worse, to write—that we expect to revolutionize a field with our work reflects either extremely youthful naiveté or access to some very good drugs.

Do not neglect first and second impressions. In other words, try to make your proposal inviting to read. Use a large font. If the agency specifies 10 point or 12 point (as most of them do), use 12. That is easier to read. Leave open spaces. Many agencies suggest single-spacing text, and if you do that, leave double spaces between paragraphs. Keep normal margins. Insert boldface occasionally, for instance, in headings, to break up the “gray” look. If the agency includes points that proposals must include (as most do), make those points easy to find in your proposal because reviewers may be asked to check them off. For instance, write “The PI will devote half-time during the summer to this project” and put that sentence at the beginning of a paragraph instead of burying it somewhere in a paragraph that primarily deals with a different topic.

## **IX WHAT HAPPENS NEXT?**

After you have produced the best proposal you can, using all the help and advice you can persuade colleagues and others to provide, you turn over the file to a research officer who submits it electronically or prints it and sends copies to the agency through the U.S. Postal Service. Knowing what happens next reduces anomie during the waiting period that comes next. It may also affect the way you prepare and write the proposal in the first place.

Federal agencies and large foundations rely on external reviewers—the readers mentioned in many places in this chapter. Smaller foundations and smaller governments usually ask one of their in-house program officers to review proposals. Naturally, the more people who will review your proposal, the longer the process will take. For federal proposals, you should expect 6 months to elapse between submission and decision, although sometimes the time is somewhat shorter.

If a funding agency uses external reviewers, where do those come from? Most of them are academic social scientists, although scientists in research firms also review some proposals. These reviewers are picked by program officers, and there are many places program officers look for them. Here are a few:

- Scientists known to the program officer, perhaps from their own earlier proposals or from the program officer's knowledge of the discipline and files maintained by the agency.
- Publications in professional journals and papers read at professional societies. The program of a society's recent annual meetings is an excellent source of potential reviewers who understand a topic area.
- Citations in the proposal. If a PI cites someone's work as part of the background to the proposed research, that generally endorses that person as a reviewer. Notice whom you cite in the first few pages of the proposal because the program officer is not going to send it to everyone you cite, and you may wish to make certain that some particularly relevant candidates do get asked.
- Suggestions from the PI. Yes, most agencies allow PIs to suggest potential reviewers, and some even permit requesting some names that should not be used. It is good to use this option when it is available, but a PI who does it without good faith does himself or herself serious damage (suggesting your mother to review is a mistake). Program officers can spot unfair suggestions, and they have long memories of PIs who tried to trick them.

There are other sources of reviewers, but these four give a fair range of the places they come from. I hope they also help explain some of the earlier suggestions about proposal preparation. A focused literature review, for instance, not only places the proposed work in intellectual context. It also suggests some of the scholars who might be suitable reviewers for it, for they know better than anyone else whether a proposal makes appropriate use of their work.

When writing, it is helpful to try taking the role of the other, as George Herbert Mead might say. That is, try to imagine a reviewer reading your proposal. We already considered what makes a worthwhile topic—namely, one related to topics others have investigated. This means there is a good chance reviewers will already have an interest in finding out what the proposed research will show. The PI does not, as noted previously, need to spend a great deal of effort convincing them it is worthwhile.

Now consider what this part of the life of a reviewer might be like. Members of NSF panels and NIH Study Sections may have a couple dozen or more proposals to read, think about, and then write reviews of. These reviewers usually have approximately a month to do those chores, often less. I hope that helps explain the significance of clear, vivid writing.

We can go further. Imagine a reviewer with a stack of proposals. In the evening, after the dishwasher is loaded and running and the kids are put to bed, he or she will sit down with the stack and eventually read yours: *one time*. That is right: he or she will probably only read it once. When I have given talks on

proposal writing, I always tell people this, and I always dread the strong negative expressions that inevitably follow. “You mean I could spend half a year of my life writing a proposal and some SOB is only going to read it once?” Yes; welcome to the world.

If a proposal stands out from the pack enough that a reviewer considers recommending funding, then he or she might go back to it later. But the first impression comes from the first quick reading, and that has to be a good impression. The ideas should leap off the page and excite the reader. Leaving aside your own emotional reactions and “shoulds,” knowing this may provide motivation to put a little more time on a proposal. Make it clear, write for good flow, make it easy to review key points, and do everything else so that your proposal gets the most out of the one-time pass from a reviewer.

While your proposal is under review, you are free to communicate with the program officer to find out where it is in the process, whether you should provide additional information, or for any other reason. Most of the time, program officers welcome such contacts. If circumstances make him or her too busy when you call or e-mail, you will get a return message soon. Program officers are dedicated people, and many of them have been on the other side writing proposals. They are sympathetic. Most program officers view their task as representing researchers to the funding agency, not the other way around. They really consider themselves your representatives. They know writing a proposal is hard work and that it has high stakes for the PI. They are much more welcoming, helpful, and friendly than many PIs presume, so any contact you initiate is likely to lead to a happy surprise.

## X SUCCESSFUL AND UNSUCCESSFUL PROPOSALS

If your proposal succeeds, congratulations. That means you will have the opportunity to do the work, with intellectual satisfaction and possibly career enhancement. Tell people. Your immediate supervisor at work (your department chair or dean if you are a university faculty member) will want to know. You might point out that fewer than half the proposals considered receive funding, and you could further point out that fewer than half the social scientists even ever write a proposal. A successful proposal moves you to the top of your class.

What about the less happy situation of a declined proposal? Most scholars I know read quickly through decline notifications (from journals as well as from funding agencies) and then put them aside for a few days. It is not pleasant having one’s work rejected, after all. When you feel ready to get back to it, you might begin by looking on the bright side. First, look at the benefits you have received, even though funding is not one of them just yet. Writing a proposal clarifies your ideas, and it is as much an intellectual product as is a research paper. Also, by submitting a proposal, you receive a great deal of feedback from very competent people on your work. Reviewers’ comments and often also the thoughts of a program officer will come your way. In a world in which most of

one's immediate colleagues do not have time to think about each other's work or ideas, that feedback is invaluable.

Treat the decline as a "revise and resubmit" if you think you can do better next time. Take the reviews seriously, think about them, and try to respond to them in revising your proposal. Incorporate changes to address what reviewers saw as weaknesses, and play up the parts they saw as strengths. Try again. People who succeed, persevere. The people who are most often funded are the same ones who are most often declined. They are the ones who do not give up. A sure route to failure is to submit a proposal and if it is declined to conclude that "I guess I don't know how to do this." Nonsense! Nobody knows for sure how to do it. As with most things in life, a wise person makes attempts, watches the results, tries to do better next time based on them, and does not give up.

Bob Lucas, a research professional who writes advice on research and proposal writing, wrote, "When all is said and done, the best way to get a grant is to write a proposal." My observation is: "They seldom send you money that you haven't asked for." Try! And good luck to you.

## ACKNOWLEDGMENTS

Lesley A. Brown, Director of Proposal Development at UNC–Charlotte, provided much useful information and many helpful suggestions for this chapter. The chapter's weaknesses are mine.

## FURTHER READING

- Chapin, P. G. (2004). *Research projects and research proposals: A guide for scientists seeking funding*. New York: Cambridge University Press.
- Friedland, A. J., & Folt, C. L. (2000). *Writing successful science proposals*. New Haven, CT: Yale University Press.
- Ogden, T. E., & Goldberg, I. A. (2002). *Research proposals: A guide to success* (3rd ed.). San Diego: Academic Press.

# Index

Note: Page numbers followed by *f* indicate figures and *t* indicate tables.

## A

- Abstract considerations, 127
  - expectations states, 133
  - manipulations, 134–135
  - manipulations checks, 135
  - standard protocols, 133
  - status information, 133
  - variables and conditions, 133–134
- Agenda mechanisms; *see also* economics and political science
  - “backward-moving”, 326
  - group decisions, 326, 327*f*
  - institutional mechanisms, 325–326
  - natural setting, 326
  - spatial committee experiments, 327
- Applied experimental research
  - active and substantial contribution, 469
  - actual end-user organizations, 468
  - vs.* basic research, 455–456, 464
  - characteristic, 464
  - computer-mediated communication system, 465–466
  - content areas, 452
  - crew resource management training, 469–470
  - ethical principles, 467–468
  - government approval, 469
  - historical period, 452–455
  - instantiate theory, 464
  - laboratory procedures, 466–467
  - practical issues, 452
  - practice, 464
  - proposal and generating funding, 457–464
  - publications, 468–469
  - recruiting research participants, 466
  - role of theory, 456–457
  - sponsor-requested presentations, 468
  - technical communications, 465
  - test theory, 464
  - theoretical and practical value, 469
  - theoretical research program, 470
  - types, 465
  - user community, 465

- Assistant training, 175–176
  - double-blind experiments, 89–90
  - expectation states, 88
  - materials selection, 88
  - novice experimenter, 89
  - planned reading program, 88
  - reading list, 89
  - reading program creation, 88
  - scope conditions, 88
  - skills and responsibilities, 90

## B

- Bales’ effect, 185–187
- Bales Interaction Recorders (1950), 91
- Brunswik’s lens model, 405–406, 406*f*

## C

- Cause–effect, 54, 57–58, 59
  - media scholars, 397–398
  - relationships, 386
- Coding interaction data
  - Bales Interaction Recorders (1950), 91
  - interaction coders, 91
  - reliability assessment, 92–93
  - time stamp, 166
  - training, coder, 92
  - transcript preparation, 92
- Compensation of participants, 43–44, 45, 94
- Confederates, experimental, 35, 137, 183
  - individuality and memorability, 156
  - management and assessment, 157–158
  - observation and listening, 158
  - post-task questionnaire, 158
  - training, 157
- Consistency, 26, 45–46, 58, 79, 211–212, 237–238, 258
- Contiguity, events, 57, 64
- Coordination failure, 433–434
  - design and implementation
    - buyer and seller, 364–365
    - experimental economics, 364

- Coordination failure (*Continued*)  
 financial incentives, 365  
 gift-exchange games, 365  
 interaction protocol, 367  
 minimum and chosen effort, 365  
 “one-shot” games, 367  
 participants’ decisions, 366  
 payoff-improving vs. safe action choices, 366  
 stimuli materials, 366–367  
 efficiency-enhancing design choice, 379  
 experimental protocols, 357  
 fixed matching protocols, 379–380  
 games; *see* coordination games  
 micro- and macroeconomics, 357  
 MouseLab/eye-tracking technologies, 380  
 real-life markets, 380–381  
 technological developments, 367–368
- Coordination games  
 critical-mass, 362–364, 377–378  
 mixed-motives, 361–362  
 Pareto-ranked, 359–361  
 pure/Rendezvous games, 358–359
- Covariation, 56–57
- Crew resource management (CRM) training, 469–470
- D**
- Data, analysis and interpretation  
 experimenter effects  
   decision-making model, 142  
   dependent variable, data collection and measurement, 143  
   double-blind, social science experiments, 140–141  
   double check, 141  
   expectancy, 141  
   experimenter–participant contact, 141–142, 143  
   natural variation, 141  
   observer/interpreter effects, 141  
   paid volunteer participants, 142  
   reliability, 141  
 power analyses, 492  
   dependent variable, 139–140  
   online calculator, 140  
   participants number determination, 139  
   sample size determination, 140  
   Type I error, 139  
   Type II error reduction, 139
- Deception, 85, 104, 118, 148–154  
 active, 35
- benefits, 34–35  
 checklist, 35  
 decision-making ability, 33  
 definition, 33  
 dehoaxing, 35–36  
 desensitizing, 35–36  
 direct harm to subjects, 34  
 economists, 33–34  
 future investigations and, 34  
 harmfulness, 35  
 human dignity in society, 34  
 potential harms, 33  
 social scientists debate, 36
- Decision makers’ (DMs’), 98, 462 *see also*  
 judgment and decision-making (JDM)  
 research  
 Bayes’ theorem, 405  
 contrast normative and descriptive models, 407  
 decision strategies, 417  
 EII theory, 409  
 evacuation, 404–405  
 EV model, 414  
 experience adverse effects, 412  
 FTT, 415–416  
 hindsight bias, 410–411  
 JDM heuristics, 423  
 judgment, 405  
 judgmental accuracy, 405–406  
 persuasion, 420–421  
 prospect theory proposes, 412  
 replacement, 424  
 semantic memory, 418  
 subjective probability judgments, 409, 410  
 working memory, 419
- Defense Technical Information Center’s Scientific and Technical Information Network (STINET), 459, 468–469
- Department of Defense (DOD), 458–459, 466
- Dependent variables, 6–7, 266, 278, 305, 386, 492
- Designing experiment, 151, 153  
 double-blind experiments, 71, 89–90  
 hypotheses, 130  
 manipulations, 130–132  
 network exchange theory (NET), 128  
 standard protocols, 129  
 theoretical assumptions, 128  
 understandable and relevant, 129  
 variables and conditions, 129–130
- Double auction game  
 asset markets, 342

- classroom experiment, 340  
 competitive equilibrium, 340–341  
 institutions matter, 341  
 research institutions comparing, 341–342
- E**
- Economic games, 298  
 buyers and sellers, 339–340  
 Canonical Games I, 340–342  
 Canonical Games II, 342–344  
 demand and supply, 340  
 ESA, 335  
 fairness and bounded rationality, 336  
 gender differences, 350  
 “Golden Fleece” award, 337  
 heterogeneity, 350  
 instructions and protocols, 339  
 markets, 335  
 microwave spectrum auction, 336  
 NSF, 337  
 payment, 339  
 public goods/bargaining games, 338  
 self-reported survey, 349  
 substantial monetary stakes, 338  
 The trust game, 348  
 The ultimatum game, 345–347  
 Economics and political science, 241  
 agendas, 325–327  
 asymmetric relations, 327–329  
 canonical experiment  
     actors, 313  
     Arrow, Kenneth, 312  
     decision rule, 311–312  
     democratic principles, 312  
     “induced valuation”, 315–316  
     payoff functions, 314–315, 315*t*  
     policy space, 314  
     proposal, 318–319  
     sample decision screen, 318, 318*f*  
     spatial committee experiments, 316–317,  
     316*f*  
     spatial model, 312–313  
     voting and agenda rules, 317
- equilibrium  
     amending and adjourning, votes, 320  
     baseline core trials, 320–321, 321*t*  
     group decisions, 319–320, 320*f*  
     proposals, 321
- nonequilibrium  
     baseline skew star trials, 324–325, 325*t*  
     descriptive data, 322–323, 323*t*
- group decisions, 321–322, 322*f*, 324*f*  
 pentagon, 322  
 star and skew star distribution, 321–322  
 star configuration, 324  
 redux, equilibrium and disequilibrium,  
     329–330  
 “standard experiment”, 311
- EII theory; *see* explicit–implicit interaction (EII) theory
- Ethics and experiments  
 IRB; *see* institutional review boards (IRBs)  
 in laboratory experiments, 27–38  
 Nuremberg Code, 23  
 regulatory requirements; *see* regulatory requirements  
 in research, 85  
     Belmont Report, 27  
     beneficence, 26–27  
     definition, 25  
     human dignity, 26–27  
     justice, 26–27  
     Kelman, Herbert, 26  
     persons respect, 26–27  
     plagiarism, 25  
     scientific misconduct discussions, 25–26  
     shortcuts, 25  
     society benefit and research participants  
     rights, 25  
     social scientists, 23
- Expectation states theory; *see* standardized experimental situation (SES)
- Expected utility (EU) principle, 406, 407,  
     422–423
- Expected value (EV) principle, 406, 407–409,  
     414, 422–423
- Experimental manipulations and deception,  
     43–44, 299–300  
 audio connection, verbal interaction, 152  
 checks, 151–152  
 confederates, 152  
 design issues, 145  
     length of the experiment, 154  
     participant involvement, 155  
     stress and discomfort, 155
- ethical codes, 149
- experimental records, 149
- informed consent, 149
- intentional misrepresentation, 148–149
- laboratory experiments in sociology, 153  
 network, 153–154
- organizational structure, 153–154
- payoff structure, 151–152

- Experimental manipulations and deception  
*(Continued)*
- “personal response style” type, 150
  - social psychological studies, 153
  - status-construction, 150, 151
  - structure interaction, 150
  - subordination, dyad, 153–154
  - treatment and outcome decisions, 151
  - written/verbal instructions, 150
- Experimental research, 10, 14–19, 469
- confederates, 156–158
  - dependent variable measurement, 145
  - design issues, 154–155
  - emotional responses, 145–146
  - equipment and facilities, 147
  - experimental treatment, 146
  - experimenter effects, 175
  - forms and documents, 148
  - information technologies staff, 147
  - institutional review boards, 147–148
  - lab facility, 145–146
  - laboratory space, 146
  - manipulations and deception,
    - experimental, 148–154
  - participant confidentiality, 147–148
  - payment and credit issues, 173–175
  - periodic checks, 147
  - postexperimental debriefing, 146–147
  - pretesting, 160–162
  - procedure development, 158–160
  - research legitimacy, 147
  - risk subjects’ task orientation, 146
  - status-construction experiment,
    - 146–147
  - subject pool maintenance, 168–173
  - tools and toolbox, 147
  - video recordings, 162–168
- Experimental staff, 143
- cash inducement, 95
  - communication
    - “conversations”, 97
    - encourage participation, 97
    - meetings, 97
  - group identity, 96–97
  - punctuality lessons, 95
  - social cohesion, 96–97
  - socialization, 95
- Experiments
- advantages and disadvantages, 91
  - abstract and, 13
  - direct testing, 10–11
  - dyads cooperation, 14
  - experiments permit direct comparisons, 10–11
  - independent variables, 10
  - natural setting, 11–12
  - public goods, 12–13
  - random assignment, 11
  - temporal/theoretical ordering, 12
  - well-designed experiment, 10–11
- development
- abstract considerations; *see abstract considerations*
  - data, analysis and interpretation; *see data, analysis and interpretation*
  - designing; *see designing experiment*
  - pretesting and pilot testing, 135–138
- history, 183
- continuing growth and development, 6
  - designs and “tastes” in design, 8
  - new technology, 5–6
  - Pavlov’s training, 7–8
  - research design, 6–7
  - respondents, random allocation, 7
  - techniques, 8
  - topics and theory, 5
  - training, 6
- social science research, designs comparisons; *see social science research*
- steps
- abstract design, 14
  - confirmation and disconfirmation, 18–19
  - experimental operations, 16
  - foundations, 14–15
  - interactants, 16
  - interpretation, 14
  - introductions and instructions, 17
  - limits, 17
  - operations, 14
  - scope conditions, 14–15
  - status characteristics, 16–17
  - “task-focused”, 17
- as theater, 421
- “backstage” and “frontstage” behavior, 86–87
  - dramatic event, 85–86
  - emotional tone and impressions, management, 86
  - experienced assistants, 87
  - formal notes, 87
  - legitimate researcher, 87
  - props, 87
  - rehearsals, 87
  - research assistants, 86

- script development, 87  
 theatrical directors, 87  
 training program, design, 86
- Explicit–implicit interaction (EII) theory, 409
- External validity and artificiality, 109–110, 192–195  
 definition, 77  
 empirically driven experiments, 78–79  
 statistical inference, 77  
 theory-driven experiments, 77–78
- F**
- Fast and frugal heuristics (FFH)  
 approach, 415–416
- Field experiments, 121, 296–297, 298  
 comparative politics, 298  
 political attitudes and beliefs, 298  
 subfields, 296  
 voting behavior, 299
- Funding, 174, 457–464  
 description, 479  
 funding agencies, 479–480  
 institutional structures, 479  
 legal requirements, 479  
 program officers; *see* program officer researcher, 481
- Fuzzy-trace theory (FTT), 415–416
- G**
- “The Gender experiment”; *see* abstract considerations
- Graph-theoretic power index (GPI), 204
- Group decision-making  
 agenda manipulations, 327*f*  
 democratic fairness, 312  
 disequilibrium “star” preferences, 322*f*, 324*f*  
 equilibrium, 320*f*  
 “standard experiment”, 311
- H**
- Holy triangle, 280–281
- Human participants, in laboratory experiments  
 “convincingness”, 104  
 critic, 104  
 experimenters counter, 104  
 “group processes”, 103  
 “informed consent”, 104  
 institutional review boards (IRBs), 104  
 in political science, 103  
 field, 121–122
- laboratory locations, 120  
 nonstudent participants, 120–121  
 survey and online experiments, 122–123
- in psychology, 103  
 ethical considerations, 106–109  
 lessons, 113  
 methodological considerations, 109–111  
 participant pool management, 111–113  
 solicitation, 105–106
- in sociology, 103  
 online experimentation and virtual worlds, 118–119  
 recruitment methods, 116–118  
 selection criteria, 115–116  
 volunteer-required participation, 115
- undergraduate education, 104
- I**
- Independent variables, 6–7
- Institutional review boards (IRBs), 491  
 advancing productive interactions  
 adversarial relationship, 46  
 human dignity, maintenance, 47  
 human research participants, treatment, 46  
 IRB-researcher interactions, 47  
 IRB staff and committee members, 47  
 moral thinking, 46–47  
 mutual education, 46–47  
 obstructionists, 46  
 people’s rights, 47  
 procedural and interactional justice, 47  
 role-taking, 46–47  
 scholars, future generations, 46–47  
 symbolic interactionist concept, 46–47
- application materials, 43*t*  
 consent procedures, 40–41  
 exclusions, 41  
 “exempt” review category, 41–42  
 federal regulations, 41  
 guidelines, 24  
 guiding researchers, 41  
 institutions, 41  
 minimal risk studies, 41–42  
 OHRP, federal regulations and advisory postings, 40  
 proposal preparation, 491  
 application, 42  
 application materials, 43*t*  
 compensation, 45  
 federal regulations, 45  
 flexibility inherent, 44–45

- Institutional review boards (IRBs) (*Continued*)  
 informed consent process, 44  
 investigators, 46  
 language, 42  
 materials preparation, 42  
 participants characteristics, 42–43  
 protocol, 42–43  
 reviewer's concerns, 45–46  
 risks, researcher's analysis, 43–44  
 researchers' responsibilities, research  
     participants protection, 38, 40–41, 40*t*  
 social science research  
     biomedical model of research, 39  
     direct benefits, 39  
     failures, 39–40  
     human dignity, 38–39  
     qualitatively trained sociologist, direct  
         observation, 39  
     scholars, 39–40  
     social scientists, vocal critics, 39
- Interaction coders, 91
- Internal validity, 192, 297  
 designs and threats, 313  
 confound, 68  
 experimental mortality, 70  
 experimenter bias, 70  
 history, 69  
 instrumentation, 70  
 maturation, 69  
 “one-shot post hoc” experiment, 68  
 regression, 69–70  
 selection-maturation interaction, 69  
 testing, 69
- experimental design to resolve problems  
 control group, 71  
 double-blind experiment, 71  
 factorial designs, 72–73, 73*f*  
 quasi-control group, 71–72  
 randomly assigned, 72  
 single-blind experiment, 71  
 two-factor design, 72
- J**
- Judgment and decision-making (JDM) research  
 anticipatory emotion, 423  
 Bayes' theorem, 405  
 Brunswik's lens model, 405–406, 406*f*  
 cognitive constraints  
     attention, 418–419  
     bounded rationality, 417  
     laboratory research, 417  
     long-term memory, 417–418
- minimize decision effort, 417  
 working memory, 419
- decision alternatives, 403–404, 408–409  
 description, 403  
 disadvantages, 406–407  
 distinct experiential and analytic  
     systems, 423
- and DMs, 403–404  
 dual-system models, 420–421  
 and emotions, 421–422  
 evacuation, 404–405  
 false-positive/negative error, 404–405  
 frequency distribution, 407  
 heuristics  
     availability, 414  
     and biases, 403  
     definition, 413–414  
     EV model, 414  
     FFH approach, 415  
     FTT, 415–416  
     interpretations, 414  
     probabilities deleted version, 415–416  
     representativeness, 413–414  
     researchers, 415
- MCPL, 405
- medical diagnosis and personnel selection,  
 425
- normative and descriptive models, 407–408,  
 424
- principle, 404–405
- probabilities, 404
- Pythagorean theorem, 422–423
- random errors, 424
- rational decision-making, 406
- regression analyses, 405–406
- search for additional information, 412–413
- SEU, 407
- subjective probability judgments, 409–411
- task constraints, 416–417
- tentative efforts, 424–425
- unobservable state, 405
- utility judgments, 411–412
- weather forecasters, 416
- L**
- Laboratory experiments, 14, 212  
 ethical concerns  
     coercive, exploitative and intrusive  
         practices, 30–31  
     deception, 33–36  
     objectification, 27–28  
     potential harms, 28–30

- privacy and confidentiality, maintenance, 31–32  
subject pools, 37–38
- Laboratory experiments, sociology assessment  
conceptual and methodological tools, 190  
effect experiments and programs, 191  
instantiation and scope conditions, 191  
network, theories and models, 190  
theoretical research programs, 190–191  
autokinetic experiment, 184  
effect research programs, 184–185  
external validity  
design and analysis, 194  
effect-oriented experiments, 192–193  
errors, instantiation, 194  
Hawthorne effect, 194–195  
randomization, 192  
simulation, 194–195  
source diagnosis, 193–194  
testing and remedying, 194–195  
testing theories, 193  
group decision, 184  
Lewin's group decision, 184  
theoretical research programs, 188–190
- M**
- Manipulations, 29, 62  
abstract considerations, 134–135  
behaviors and self-reports, 265–266  
designing experiment, 130–132  
exclusions, 264  
experimental error, 262–263  
independent variables, 263  
irrelevant variables, theoretical interest, 264  
operationalizations; *see* operationalizations and manipulations  
scope conditions, 263  
test limitations, 264
- Marginal per capita return (MPCR), 342–343
- Media stereotyping effects  
African-American targets, 388–389  
audience agency, 396  
audiences outcomes, 386  
cause–effect inferences, 397–398  
computer-based tools, 397  
context of “traditional” media formats, 396  
counter-stereotypes, parasocial contact and stereotype change, 390–391  
cultivation and mental models, 386, 387–388  
dependent measures, 394–396  
description, 386
- independent variables, 392–393  
individuals' attitudes, perceptions and behaviors, 385–386  
lab-based experimental research, 385, 386  
mass media, 387  
neo-associationistic model, 388  
new media, theoretical and methodological concerns, 396  
prejudice, 386–387  
priming, 388  
psychology, 385  
racial/ethnic stereotypes, the U.S. contexts, 397  
research in a post-feminist, post-racial era, 397  
sample populations, 391–392  
social identity and cognitive theory, 389–390  
standard designs and procedures, 391  
stimuli based, 393–394  
theoretical implications, 397
- Mill's canons and inferring causality, 20  
assess causation, 59  
causal factor, 61  
limitations, 61–62  
method of agreement, 59, 60f  
method of difference, 59, 60f
- Mixed-motives coordination games  
“battle of the sexes”, 361, 362, 362t experiment, 370–372
- Morphing technology, 300
- MPCR; *see* marginal per capita return (MPCR)
- Multiplayer online role-playing games (MMORPGs), 119
- Multiple cue probability learning (MCPL), 405
- N**
- National Institute of Justice (NIJ), 473–474  
National Institutes of Health (NIH), 473–474, 484, 495, 500
- National Science Foundation (NSF), 473–474, 479, 484, 495, 500
- Network exchange theory (NET), 128  
exclusionary networks, 216  
GPI, 204  
power-dependence theory, 204, 208
- Nonspuriousness, 58
- N-person social dilemmas  
incentive structure, 230  
individual-level defection, 228  
marginal per capita return (MPCR), 228–229  
one-time decision-making, 228–229

*N-person social dilemmas (Continued)*

- production function, 228
- record sheets, 229
- standard linear public goods setting, 228–229
- static public and resource goods, 229, 230
- static settings, 230
- utility formulations, 230

**O**

Objectification, 27–28

Observable power and prestige order (OPPO), 271

Office of Sponsored Programs and Research, 474

Operationalizations and manipulations  
collective orientation, laboratory, 260–261  
comprehension-related barriers, 260  
contrast sensitivity, 260  
definition, 255  
experimental limitations, 259  
irrelevant variables, 258–259  
levels of performance evaluations, 255–256  
random assignment, 255–256  
replications, 262  
sex category and skin color, 261  
stress competitiveness, 260–261  
task's sex linkage, 261  
test diagnosticity, 255–256  
theoretical variables  
dependent, 258  
independent, 257  
irrelevant, 258  
scope conditions, 256–257

Order-statistic games, 372–377

Organizational behavior, 79

characterization, 433

description, 433

ethical behavior

- accountability systems, 439
- dispositional factors, 436–437
- ethical cultures, 438–439
- hierarchical arrangements, 437
- indispensable, 439–440
- individual characteristics, 435, 436
- individuals' moral identity, 436–437
- interdependence-based problem, 434
- joint work, 434–435
- measurement, 436
- moral issue, 437–438
- participants, 435
- passive observational studies, 437–438

primes and structural manipulations, 437

problems, 434

rationalist models, 436

research, 435

self-report studies, 436

social contexts, 439–440

traditional passive observational

methods, 435

workplace, 434–435

experimental study, 434

interdependent work, 433–434, 440

intervention in employee training, 442

nonexperimental methods, 441

passive observational studies, 440–441

psychological science, 442

researchers, 442

unfairness, 433–434

**P**

Pareto-ranked equilibria, 363–364, 366–367, 372

global games, 361

minimum-effort/weak-link game, 361

order-statistic, 361

“security”, 360–361

stag-hunt game, 360, 361

“strategic uncertainty”, 359, 360t

Payment and credit issues

experimental sessions, 174–175

financial services personnel, 173–174

funding agencies, 174

research assistants, 174

social science experimentation, 173–174

Planned reading program, 88

Political science, 103, 119–123, 241, 312–313

American politics and voting

behavior, 296, 299

assessment and challenges, 301–302

attitudes and beliefs, 298

bargaining and negotiations, 298

biological and genetic influences, 304

British government, 303

campaigns and elections, 299–300

cognitive neurosciences, 302

corruption, Brazil, 301

costs and risks, 295–296

decisions and behaviors, 304

dependent and independent variables, 305

“dust bowl” empiricism, 305

ecological validity and generalization, 299

and economics; *see* economics and political science

- EEG technology, 301  
 free-rider problems, 298–299  
 gender differences, 302–303  
 genetic determinants, 304  
 genome-wide association studies, 303  
 Hispanic picture, 300–301  
 mayor’s election, 296–297  
 monetary *vs.* nonmonetary payoffs, 302  
 morphing technology, 300  
 MRI study, 306  
 online survey instrument, 300  
 psychological models, 300  
 rhetoric and communication styles, 303  
 voting behavior; *see* voting  
 WHI, 295
- Political science, participants; *see also* human participants, in laboratory experiments in field experiments, 121–122  
 laboratory locations  
   “collective decision-making”, 120  
   college dormitories visit, 120  
   Interdisciplinary Experimental Laboratory (IEL), 120  
   non-internet methods, 120  
 nonstudent participants, laboratory locations, 120–121  
 survey and online experiments, 122–123
- Potential harms, 28, 30, 33, 34, 36, 43–44  
 economic and legal harms, 29–30  
 inconvenience, 29  
 physical, 28–29  
 psychological harm, 29  
 social harms, 29  
 types and degrees, 28
- Power and equity, exchange networks  
 balance/imbalance, 214  
 Emerson’s power-dependence theory, 214  
 high-value exchange relations, 214–215  
 mean power use, 216  
 mixed design, 216  
 networks in exchange experiments, 214, 215f  
 primary dependent variable, 216
- Prerecorded instruction production, 165, 175  
 computer editing programs, 167  
 digital video technology, 167  
 mechanical/electronic hardware, 167–168  
 narrator’s presentation, 167  
 pretesting, 166–167  
 recording sessions, 166  
 scheduling practice sessions, 166  
 software programs, 166–167  
 video recordings, 166–167
- well-labeled digital file, 167  
 Pretesting, 132, 155  
   collective orientation, 160  
   contrast sensitivity task, 160  
   data collection instrumentation, 160  
   decision rules for exclusion, 162  
   disagreement trials, 160  
   independent variable manipulation, 160  
   laboratory personnel, 162  
   low task score, 161  
   manipulation checks, 160  
   paper-and-pencil instrument, 160  
   and pilot testing, 135, 155  
     abstract and theoretical concerns, 137  
     attention, 137  
     competing processes, 138  
     cultural factors, 136  
     design elements, examination, 135–136  
     double-check systems, 136  
     informants, 135  
     multiple experimenters conduct, 136  
     participants, 136  
     rehearsals, 138  
     technology, 136  
 postsession interviews, 161  
 procedural changes, feedback, 161  
 questionnaire items, 160–161
- Privacy and confidentiality, maintenance, 31–32
- Program officer, 480, 481, 501  
 applicant, 481  
 call and e-mail, 482  
 funding agencies, 480, 481–482  
 proposals, 476  
 relationship to researchers, 480  
 reviewers, 500  
 roles, 482
- Psychology, participants; *see also* human participants, in laboratory experiments  
 ethical considerations  
   coerciveness, 106–107  
   educational value, 108–109  
   “totally volunteer” participants, 106–107
- management  
   annual fees, 113  
   experimenters, 112–113  
   instructors, 112–113  
   paper-and-pencil sign-up sheets, 112  
   site administrators and research personnel, 112–113  
   student appointments schedule, 111–112  
   Web-based experiment management systems, 112–113

- P**
 Psychology, participants (*Continued*)
  - methodological considerations
    - external validity, 109–110
    - homogeneous nonprobability samples, 110
    - measurement error source, 110
    - participant pools, homogeneity, 110
    - scope conditions, 110
    - task oriented and collectively oriented, 110
    - Type II statistical error, 111
  - “mimetic process”, 105
  - recommendations, 113, 114t
  - solicitation, 105–106
 Public goods game, 347, 348
  - Bohm, Peter, 342
  - description, behavior, 344
  - human behavior, 344
  - hypotheses, 343
  - monetary payoffs, 343–344
  - MPCR, 342–343
  - Nash equilibrium, 343
  - punishment behavior, 344
 Pure coordination games, 361, 369–370
  - labels, 359
  - matching games, 369
  - “payoff” table, 358, 358t
  - primary vs. secondary salience, 369
  - Schelling salience, 370
  - self-enforcing, 358–359
  - strategic uncertainty, 358
  - The Strategy of Conflict*, 359**R**
 Random assignment, 63
 Regulatory requirements, 23, 25
  - debate centers, 38
  - human dignity protection, 38
 IRB review; *see* institutional review boards (IRBs)
   
**S**
 SES; *see* standardized experimental situation (SES)
 SEU; *see* subjective expected utility (SEU)
 Social dilemma experiments, 151–152, 242
  - behavioral economics, 241
  - behavioral political science, 241
  - characteristics, group members, 237
  - communication, 237–239
  - economics and political science, 241
  - empirical application, 236
  - game theoretic models, 241
  - information rules, 242
 institutional rules, 241–242
 interdisciplinary research, 226
 laboratory and field experiments, 226
 nonexcludability, 241–242
 N-person social dilemmas, 227–230
 public and resource goods, 225
 punishment and triggers, 239
 simultaneous/sequential decisions, 235–236
 strong free-riding hypothesis, 230–232
 symmetrical vs. asymmetrical information, 234–235
 testing payoff properties, 232–234
 theoretical approaches, 240
 theoretical coordination, 242
 theoretical language and assumptions, 240
 two-person dilemmas, 226–227
 types, 225
 uncertainty, 240
 Social exchange and networks, 203–204, 206
  - benefits and costs, 200
  - commitment, exchange relations, 220, 221
    - cohesion and commitment, 219
    - effects, power structures, 217–218
    - equal-power conditions, 218
    - relational cohesion, 219
  - computer-mediated exchange, 212–213
  - direct, generalized and productive structures, 200, 201f
  - discrete transaction, 200–202
  - dynamic processes, power use, 204
  - The Early Theories*, 202
  - economic sphere, 199
  - The Emergence of New Theories of Power in Networks*, 203–204
  - Emerson’s Power-Dependence Theory*, 202–203
  - experimental settings, 205
  - “one-shot” transactions, 200
  - power and equity, 214–216
  - reciprocity structure, 205
  - relational cohesion, 205
  - risk and trust, 219–221
  - risk and uncertainty, 205
  - standard settings
    - characteristics, structures, 205–206
    - dependent variable measurement, 210–211
    - designs and design issues, 211–212
    - generalized exchange and productive exchange settings, 210
    - network/dyadic structure, 206–207
    - reciprocal exchange setting, 209
    - social exchange networks, 206
    - standardized laboratory settings, 206

- strong and weak power, 216–217  
 theories and experimental programs, 199  
 web-based experiments, 213
- Social science research, 24, 29, 38–40, 403, 453  
 causation  
   consistency, 58  
   “constant conjunction”, 56  
   contiguity, 57  
   covariation, 56–57  
   necessary condition, 55–56  
   nonspuriousness, 58  
   sufficient condition, 55–56  
   theoretical plausibility, 59  
   time and asymmetry, 57–58
- combining different methods, 10  
 controlled experimental research, 54  
 experiments varieties  
   experimental food scientist, 75  
   internal validity and, 74  
   “one-shot post hoc” model, 74  
   population inferences, 75–76  
   prior experimental sessions, 75  
   pseudo-control group, 75  
   Type I error, 76  
   Type II error, 76
- external validity and artificiality; *see*  
 external validity and artificiality
- Fisher’s premature burial and posthumous resurrection  
 analysis centers, 65–66  
 central limit theorem, 67  
 random assignment, 64–65, 67  
 sample size and probability, 65–66, 66/  
 unknowable causes, probability and  
   control groups, 65, 65/f
- Fisher’s solution  
 controlled measurement, 64  
 experimental strategy, 62  
 manipulation, 63–64  
 random assignment, 63  
 treatment group, 62
- historical archival research, 9
- internal validity  
 designs and threats, 68–70  
 experimental design to resolve  
   problems, 71–73
- laboratory experimentation, 54
- Mill’s canons and inferring causality; *see*  
 Mill’s canons and inferring causality
- participant observation, 9
- social scientists, 54
- structured observation, 9
- survey research, 9–10  
 unnatural experimentation, 54  
 unstructured observation, 8–9
- Sociology, participants; *see also* human participants, in laboratory experiments  
 online experimentation and virtual worlds, 118–119
- recruitment methods  
 contact information, 117–118  
 dedicated scheduler, 117–118  
 e-mailing message, 116  
 handouts, 116  
 high-enrollment courses, instructors, 116  
 in-person recruiting, 116  
 paid schedulers, 117–118  
 research participation, description, 117  
 script approval, 117–118  
 semester-commitment recruiting, 116, 118  
 selection criteria, 115–116  
 volunteer/required participation, 115
- Spatial committee experiments, 316–317, 327, 330
- equilibrium, 319
- utility functions, 316/f
- Stag-hunt games, 363–364, 378–379  
 deductive principles, 377  
 “median game”, 372, 373/  
 “minimum game”, 372, 374/  
 one-way and two-way communication, 376
- Pareto-ranked; *see* Pareto-ranked equilibria, stag-hunt
- payoff-dominant/efficient equilibrium, 372  
 “weak link” description, 372
- Standardized experimental situation (SES), 6–7, 270, 498  
 construction, 270–277
- Bales’ observations, 270  
 behavioral setting, 272–274  
 experimental tasks, 274–275  
 manipulation of expectations, 271–272  
 OPPO, 271  
 standardized scenarios, 275–277  
 status-homogeneous groups, 271  
 graph formulation, status characteristic theory, 277–278
- interpersonal processes, 269
- status characteristics theory, 286–287
- testing and developing theories, 270
- theoretical research program, 269
- types, behaviors, 269
- uses and special features  
 comparability and cumulativeness, 278  
 data relevant theories, 283

- Standardized experimental situation (SES)  
*(Continued)*
- flexibility, 281–283
  - holy triangle, 280–281
  - “reverse process”, 284
  - status characteristics theory, 284
  - strong tests, theoretical arguments, 279–280
- Status characteristics theory, 16–17, 109, 110
- metatheoretical components, 286
  - theoretical components, 286–287
  - theory-based empirical models, 287
- Stimuli, 7–8, 366–367
- media sources, 393
  - pretest participants, 393
  - priming studies, 394
  - used for, 394
  - visual aspects, 393
- Strong free-riding hypothesis, 230–232
- economic models, 230
  - empirical analysis, 230–231
  - exchange types, 231
  - experimental methodology, 231–232
  - primary prediction, 230–231
  - separation, reputation effects, 231–232
  - welfare economics, 230–231
- Strong theoretical tests, 279–280
- Subjective expected utility (SEU), 407–409, 412, 423
- Subject pool maintenance, 168
- artificial setting, 168
  - contamination, 172–173
  - deception, experiment research, 169
  - demand characteristics, 168
  - ethical concerns, 37–38
  - incentive structure, 169
  - network power, 169
  - scheduling, 169–171
    - callback feature, 171
    - scheduler’s script, 170
    - scheduler telephoning, 170
    - status-construction experiment, 171
    - status-equal groups, 169–170
- T**
- TESS; *see* Time-Sharing Experiments for Social Sciences (TESS)
- Theoretical hypotheses formulation, 247, 256, 305
- “cognitive dissonance”, 248
  - contrast sensitivity, 253
  - definition, 247
- design and predictions, 253, 254t
- design decisions, 256
  - and design issues, 249
  - empirical test, 247
  - expectation states theory, 249, 250
  - experimental test, 252
  - experimenter constant, 253
  - exploratory/hypothesis testing, 248–249
  - firm empirical support, 248
  - hypothesis-testing studies, 248–249
  - laboratory experiments, 247
  - logical skill test, 249
  - operationalizations and manipulations, 255–259
  - performance evaluation level, 251
  - “performance expectations”, 249
  - research cumulativeness, 252
  - scope conditions, 250
  - “social attribution”, 248
  - status characteristics, 249
  - task diagnosticity, 251
  - task-related knowledge, 248
  - types, particular features, 256
  - variables, 248
- Theoretical plausibility, 59
- Theoretical research programs, 133, 269, 470
- analytic deconstruction, effect, 185–186
  - Asch experiment, 185
  - Bales’ effect, 185–186
  - effect programs, 185
  - experiment orientation, 187–188
  - Lewin’s group decision experiment, 185
  - “meaning insight ability”, 186–187
  - and methodological strategies, 189
  - nonverbal interaction, 188–189
  - power-prestige order, 186
  - power-prestige theory, 188
  - status characteristics theory, 188
- Time and asymmetry, 57–58
- Time-Sharing Experiments for Social Sciences (TESS), 118–119
- Time stamp, 166
- Training interviewers and experimenters, 90–91, 95
- assistant training; *see* assistant training
  - compensation; *see* compensation
  - experimental staff, as group; *see* experimental staff
  - experiment as theater; *see* experiments
  - mentor role, 42
    - maturity and motivation, 101
    - report preparation, 100
    - skills improvement, 99
    - word processing files, 100

- postsession interviews, 90–91, 100  
 audio recordings, listening, 98  
 body language identification, 98  
 end-of-session data, 98–99  
 periodic checking, 98  
 supervised practice, 98  
 preliminaries  
   assistants recruitment, 84  
   complex open-ended interview, 84  
   computer-mediated (pseudo)  
     interaction, 84  
   decisions, 84  
   job description, 83–84  
   questionnaire and interview data, 84  
   raw material nature, 85  
   in research ethics, 85  
 specific tasks  
   analysis, 93  
   coding interaction data, 91–93  
   computer-assisted experiments, 93  
   experimental sessions, 90–91  
     “recoding” process, 93  
 The Trust game, 235–236  
 altruism, 348  
 Nash equilibrium, 348  
 “reciprocity”, 348  
 World Values Survey, 348

**U**

- The Ultimatum game, 298  
 dictator and ultimatum game, 347  
 double-anonymous protocol, 347  
 heterogeneity, behavior, 347  
 income-maximizing responder, 345  
 low- and high-stakes games, 346  
 payoff-maximizing responder, 345  
 principal–agent theory, 345  
 proposer’s offer, 346–347

**V**

- Video recordings, 83–84, 160, 167–168  
 audio and lighting, 163  
 camera position, 164  
 “contrast sensitivity”, 165  
 digital recording equipment and computers,  
   162–163  
 directional microphones, 163  
 discussion groups, 163–164  
 prerecorded instruction production, 166–167  
 recorded instructions, 165  
 research assistants’ interaction, 164  
 score category, 165

- signal switcher, 164–165  
 small testing rooms, 164–165  
 stimulus material and instructions, 162–163  
 tripods, 164  
 unique code number, 164  
 verbal qualifiers, 164  
 Voting, 11, 241–242, 296, 297, 298, 299  
   American behavior, 299  
   experimental methodology, 298  
   and multicultural group behavior, 296  
   psychological models, 300

**W**

- Writing proposal, 475, 476, 501  
 contemporary empirical research, 473  
 description, 486–487  
 emotional reactions, 501  
 experimental research, 475  
 external funding, 474  
 federal agencies and foundations, 499  
 feedback, 475  
 groupings, 485  
 independent agencies, 473–474  
 indirect benefits, 475  
 investigator-initiated projects, 475  
 knowledge, 475–476  
 NIH Study Sections, 500  
 occasional small-scale investigations, 473  
 Office of Sponsored Programs and Research,  
   474  
 personnel, 486, 493  
 preparation, 484–485  
 representatives, 501  
 research funding, 474  
 research plan, 486, 491–492  
 research programs, 482–484  
 research questions, 486, 488–491  
 reviewers, 476, 500  
 social sciences, experimental research  
   artificial, 477  
   experimentalist, 477  
   mistreat human subjects, 478  
   misunderstandings and prejudices, 478  
   researchers, 476  
   research method, 477  
     “the real world”, 477  
     well-designed, 478  
   successful and unsuccessful, 501–502  
   timetable and costs, 486, 493–494  
   topic, 485, 487–488  
   the U.S. government, 473–474  
   the U.S. Postal Service, 499

- Writing proposal (*Continued*)  
writing, 476  
writing styles, 496  
    avoid distracting points, 497  
    declarative sentences, 496  
    evade page restrictions, 496  
    idea per sentence, 495–496  
    imprecision and distractions, 496  
    justification, 494  
organization, 495  
please do not, 498–499  
“power at a distance”, 494  
proofread, 496  
proposals and PIs, 494  
reviewers, 495  
“technical writing”, 495  
title, 496–497  
use bullets and summarize, 497–498