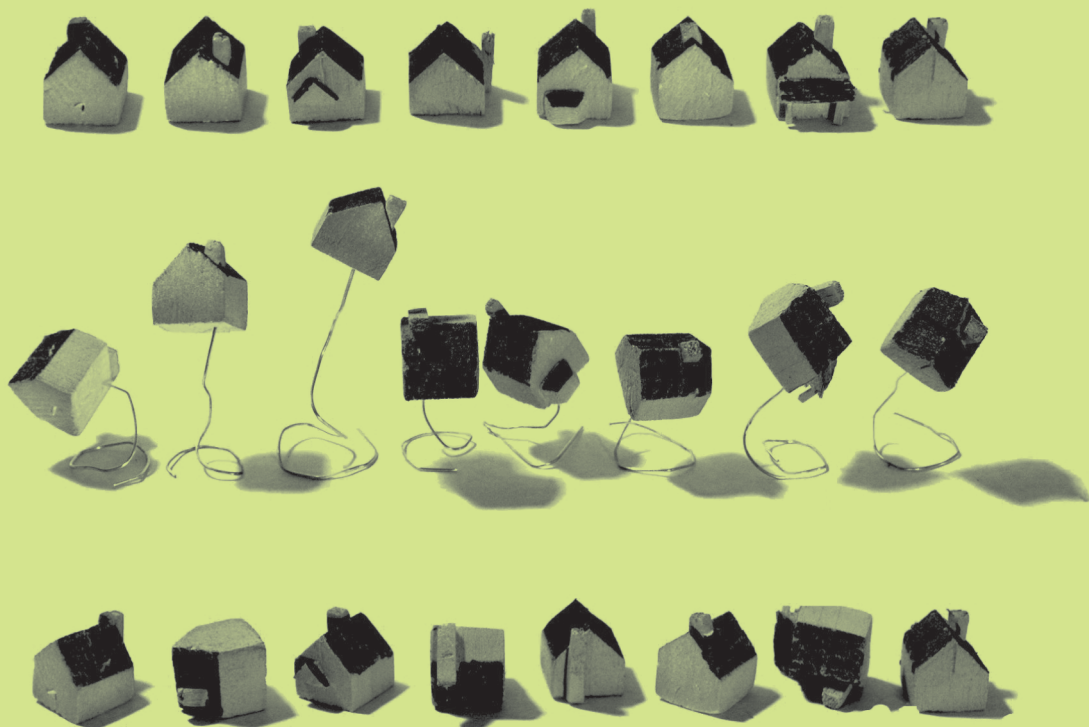


Field Experiments and Their Critics

Essays on the Uses and Abuses of
Experimentation in the Social Sciences

Edited by Dawn Langan Teele



Please do not distribute. These are Proofs.

FIELD EXPERIMENTS AND THEIR CRITICS

THE INSTITUTION FOR SOCIAL AND POLICY STUDIES AT YALE UNIVERSITY

THE YALE ISPS SERIES

FIELD EXPERIMENTS AND THEIR CRITICS

Essays on the Uses
and Abuses of
Experimentation in
the Social Sciences

Edited by
Dawn Langan Teele

Yale UNIVERSITY PRESS
New Haven & London

Copyright © 2013 by Yale University.

All rights reserved.

This book may not be reproduced, in whole or in part, including illustrations, in any form (beyond that copying permitted by Sections 107 and 108 of the U.S. Copyright Law and except by reviewers for the public press), without written permission from the publishers.

Yale University Press books may be purchased in quantity for educational, business, or promotional use. For information, please e-mail sales.press@yale.edu (U.S. office) or sales@yaleup.co.uk (U.K. office).

Set in Galliard and Copperplate 33 types by Newgen North America.
Printed in the United States of America.

Library of Congress Cataloging-in-Publication Data

Field experiments and their critics : essays on the uses and abuses of experimentation in the social sciences / edited by Dawn Langan Teele.
pages cm. — (The Yale ISPS series)

Includes bibliographical references and index.

ISBN 978-0-300-16940-9 (alk. paper)

1. Social sciences—Research—Methodology. I. Teele, Dawn Langan, 1983–

H62.F4195 2013

300.72—dc23

2013007487

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

This paper meets the requirements of ANSI/NISO Z39.48–1992
(Permanence of Paper).

10 9 8 7 6 5 4 3 2 1

CONTENTS

Preface vii

Introduction 1

DAWN LANGAN TEELE

1. The Illusion of Learning from Observational Research 000

ALAN S. GERBER, DONALD P. GREEN, AND EDWARD H. KAPLAN

2. A Defense of Observational Research 000

SUSAN C. STOKES

**3. A Retreat from Radical Skepticism: Rebalancing Theory,
Observational Data, and Randomization in Development
Economics 000**

CHRISTOPHER B. BARRETT AND MICHAEL R. CARTER

4. The Experimental Approach to Development Economics 000

ABHIJIT V. BANERJEE AND ESTHER DUFLO

5. Reflections on the Ethics of Field Experiments 000

DAWN LANGAN TEELE

**6. Instruments, Randomization, and Learning About
Development 000**

ANGUS DEATON

VI CONTENTS

7. Experimental Reasoning in Social Science 000

ANDREW GELMAN

8. Misunderstandings Between Experimentalists and Observationalists About Causal Inference 000

KOSUKE IMAI, GARY KING, AND ELIZABETH A. STUART

9. Methods Are Like People: If You Focus Only on What They Can't Do, You Will Always Be Disappointed 000

IAN SHAPIRO

References 000

Contributors 000

Index 000

PREFACE

No graduate student in the social sciences, especially not those who spent their days tromping up Hillhouse Avenue or down Prospect Street in New Haven during the early 2000s, could have missed the fierce debates incited by our experimentalists and their compatriots across the country over empirical methodology. The clever research designs of scholars at the Institute for Social and Policy Studies (ISPS), the Economic Growth Center (EGC), and Innovations for Poverty Action (IPA) at Yale were a constant topic of discussion and disagreement. There was a hopeful feeling that economists and political scientists would finally be able to address the pressing problems of the day by using a methodology that left little room for doubt. But as more was written about experimental methods, especially those carried out in the field, more questions were raised: What factors guarantee that experiments are actually implemented according to plan? How can researchers generalize from an experimental result in one locale to a policy program in another? Can experiments answer the major questions driving most social science? Are experimental protocols ethical? What, if anything, is lost by methodological monocropping?

To explore these questions I organized a debate on field experiments at Yale in October 2009. A crowd of more than 250 assembled to hear Don Green defend the experimental “juggernaut” against strong criticisms mounted by Angus Deaton, Susan Stokes, and Ian Shapiro. Along with many in the audience, I came away from the night with a sense that the debate was far from settled. This book, which contains the original essays

that inspired the debate alongside of fresh contributions, is an attempt to consider anew the arguments advocating field experiments as well as to measure and weigh criticisms of the field experimental method.

My sincerest thanks go to Ian Shapiro (of the Macmillan Center), Don Green (formerly of the ISPS), Jacob Hacker (currently at the ISPS), and Nicholas Sambanis (formerly of Yale's program in Ethics, Politics and Economics) for their support (in specie) of the original debate and for their encouragement (in spirit) for publishing this book. Thanks as well to Bill Frucht and Jaya Chatterjee at Yale University Press and to seven anonymous reviewers for their comments along the way.

Finally, I want to express gratitude to the many people who reviewed parts of this book, especially Elisabeth Wood, Susan Stokes, Frances Rosenbluth, James Robinson, Chris Udry, Matt Kocher, Casiano Hacker-Cordón, Rory Truex, Blake Emerson, and Allison Sovey Carnegie. A special shout-out is owed to my writing-group ladies, Anna Jurkevics, Erin Pineda, and Kristin Plys, with the hope that this is only the first of many times we thank one another in actual ink. Love and gratitude to Joshua Simon for finding time to read all that I write, even when he has so much to read and write himself. Finally, in memory of my grandfather Donald H. Jones, who, despite being fully informed, never stopped volunteering for randomized trials.

INTRODUCTION

Dawn Langan Teele

A central feature of the social sciences is a keen interest in causation. Among social scientists there is widespread agreement that systematic study of the social world, of institutions, economic behavior, and social action, can lead to insights about causal relationships. This agreement is not matched, however, by consensus within or across disciplines as to which research methods are most likely to achieve that goal. Part of the disagreement stems from the inherent difficulties of studying human society: in the real world, structure, intention, and accident all interact to produce complex human behaviors whose causes can be opaque, even in hindsight. To assess the relative merits of alternative theories of human behavior we need a methodology—a means of linking evidence to statements about causation.

John Stuart Mill's early reflections on these issues remain relevant today. Distinguishing between catalysts (the "causes of effects") and outcomes ("the effects of causes"), Mill thought that the key to understanding causation lay in isolating either the causes or the effects and seeing what happens. He writes, "We must be able to meet with some of the antecedents [the causes] apart from the rest, and observe what follows from them; or some of the consequents [the effects], and observe by what they are preceded."¹ In other words, we must *vary the circumstances* of the object of study to see whether a cause always has the same effect when placed in a new context, or whether effects can be traced to the same cause in different situations.²

Two main approaches have been used since Mill's time to study causation. The first is *observational* inquiry. Here, in order to pinpoint causality, the researcher looks for natural differences across cases and tries to find a single input that might have caused the variation in outcomes.³ A second technique is *experimental* inquiry. To pinpoint causality in this type of work, the researcher conducts an experiment whereby one group, the treatment group, is given a certain input that is withheld from the control group. If outcomes vary across the treatment and control groups, the researcher can argue that the difference must be due to the catalyst that she set into motion.

Mill argued that the two ways of assessing causation—experimentation and observation—were logically equivalent. But with observational inquiry he worried about the possibility that some other factor, unknown to the researcher, is actually the cause of the observed outcome. This challenges the *internal validity* of a causal inference: it is not enough to observe that y always follows x to infer that y was caused by x . As the oft-repeated phrase goes, correlation does not imply causation. In theory, random assignment in a controlled trial assures the internal validity of causal claims. In other words, if the researcher herself puts subjects into the treatment and control groups, and if she induces the catalyst whose outcome is of interest, she can say with certainty whether x does, or does not, cause y .

The insight that causal inferences must be drawn from internally valid studies was first incorporated into the natural sciences, where laboratory scientists have long analyzed treatment effects in subpopulations of bacteria and animals. In clinical medicine, randomized controlled trials are a mainstay in tracking disease progression in response to new therapies and pharmaceutical drugs.⁴ These practices have been imported into the social sciences more recently, particularly in psychology, where, as the advertisements on any student union bulletin board will attest, college students are favorite subjects in behavioral research. But there are many reasons to suspect that college students are not representative of the population at large, which raises the problem of *external validity* in controlled experiments.⁵ Because social scientists want to identify causal processes in society as a whole, the population that is studied and the conditions under which the study is carried out must be realistic enough to make the results

applicable in nonexperimental settings. Thus even randomized experiments are not immune to criticism.

In recent decades a growing cohort of researchers in the social sciences, especially economists and political scientists, have sought to gain the advantages of internal validity under experimental controls while avoiding the difficulties of external validity by conducting *field experiments*—randomized controlled trials carried out in a real-world setting.⁶ The idea is to randomly assign research participants from the real world to treatment and control groups and then intervene only in the treatment group in order to see whether the expected change actually occurs.⁷ Though Mill would have been skeptical of this move—he worried that the social world is too complicated for experimental work—the modern proponents of field experiments disagree, seeing randomized controlled studies as the missing key to sound causal inference in social research.

Yet the experimental insurrection remains incomplete. If human error or cunning leads to noncompliance with experimental prescriptions—that is, if experimental subjects do not do as they are told—the internal validity of field experiments is challenged; moreover, if field experiments are carried out on groups of people that are quite unlike the rest of us, the external validity of field experiments is suspect. In addition to those who are skeptical that field experiments can overcome these problems, there are scholars who argue that field experiments only tell us about *average* effects when in fact what is needed to test a hypothesis or make a policy recommendation is more fine-grained knowledge. A still different group of critics worries about limiting what we study to questions that lend themselves to field experimental research, which may exclude many of the most pressing issues that concern social scientists. Add to these worries the ethical concerns that arise when people are assigned to treatment and control groups without their knowledge, and it becomes clear that the debate about field experiments is far from over. The contributors to this book do not claim to end this debate, but they do offer a guide to its frontiers that will be of interest to participants and newcomers alike.

In “The Illusion of Learning from Observational Research” Alan Gerber, Donald Green, and Edward Kaplan (GGK hereafter) argue that the only upshot of nonexperimental research is its ability to teach us, when placed side by side with experimental research, how biased nonexperimental

work actually is (chapter 1). GGK develop this contentious claim by using a Bayesian framework wherein research is valued according to how dramatically it shifts our prior beliefs about a causal relationship. For them, nonexperimental work has little ability to shift prior beliefs and should, *a fortiori*, be dropped from the methodological toolbox.

Susan Stokes rejects this logic (chapter 2). She diagnoses the criticisms raised against observational research by GGK and others as part of the worldview of a “radical skeptic”—someone for whom the confluence of events that produce social outcomes is so complex that they can never fully be understood. When confronted with observational work, the radical skeptic is prone to fixate on the potential for omitted variables to sully conclusions, regardless of how sound the theory or how careful the researcher. Stokes argues that such skepticism, if applied evenhandedly to experimental research, would raise similar criticisms of experiments. The radical skeptic would worry, in particular, that heterogeneity among research participants in experimental work complicates interpretations of the average treatment effect. For Stokes, the point could be stated thus: experimentalists should clean their glasses when they read their own work; evenhanded criticism reveals that all methodologies are problematic. Unveiling truths will require more open-minded self-reflection on the part of researchers.

The economists Christopher Barrett and Michael Carter (chapter 3) build on Stokes’s framework to address the seemingly infinite ways in which the experimental ideal is violated in practice. Even if we believe that a field experiment would, if properly carried out, get us closer to the truth, in the messy practice of field research some of the assumptions that experiments rely on for sound causal inferences may be violated. For example, in a large-scale experiment a researcher may be unable to verify that the randomization of participants to treatment and control groups was properly carried out. This might happen if an NGO partner doesn’t follow instructions, or if someone on the inside thinks some subjects are more deserving or more likely to benefit from receiving the treatment. Further, a researcher who is not on-site might not be able to monitor crossover, that is, people who were assigned to one group but placed themselves in another. Both of these potential problems—imperfect randomization and participant crossover—will bias causal inferences drawn from the field experiment.

Returning to the experimentalists, Abhijit Banerjee and Esther Duflo (chapter 4), prominent development economists at the forefront of the experimental revolution in their field, disregard critiques of experiments because they apply to observational work as well. They review findings from more than ten years of experimental development economics and conclude that, despite the short timeline, experiments have produced hard facts that years of observational research cannot contend with. They argue, moreover, that the common criticism that experiments are unable to answer deep theoretical questions is unfounded: as experiments proliferate in the discipline, scholars, by refining and replicating experiments more carefully suited to the question at hand, will be able to use experimental results to refine theoretical insights.

Departing from purely methodological concerns, I raise several ethical questions about experimental research in chapter 5. When contemplating whether a given methodology is ethical, we must probe the relationship between the researcher and the research subjects. I argue that the crucial distinction between field experiments and observational research is that field experiments, by their very nature, manipulate the real world in the service of research questions. To put this another way, if observational social scientists are spectators of a card game whose hands nature dealt, experimental social scientists have positioned themselves as the dealer. This shift from spectator to dealer changes the relationship of the social scientists to the players, and it begs for an examination of the practices and policies that bind them together. I call for greater attention to be paid to the concerns and needs of study participants and for a no-exceptions policy to individual informed consent in experiments. In closing, I argue for creative thinking in research design, such as the use of placebo groups, to ensure that the spirit of the *Belmont Report*, a foundational document in research ethics, is upheld in all field experimental research.

Chapter 6 reprints, in full, Angus Deaton's "Instruments, Randomization, and Learning About Development," a much-discussed critique of experiments that inspired several of the contributions in this book. Deaton's arguments are many and complex, but they center around two main points. Following Heckman and Smith (1995), Deaton worries that the experimentalist research agenda is too focused on *what* works to the neglect of *why*. An unqualified turn toward field experiments, by this light,

would reward scholarship that focuses on small, answerable questions to the neglect of bigger, more profound concerns.

Second, Deaton argues that the “average treatment effect”—that parameter that can be reliably estimated through an experimental design—represents only a small part of what a researcher might want to know. In an experiment that offers cash transfers to the poor in exchange for putting their children in school, we might want to know not only how the average student fared in terms of educational attainment, but also other aspects of the distribution, like the median and the mode. We also might be interested in how the treatment influenced different subgroups of the population, like households with female heads or those that have many children. Deaton reminds us that in order to estimate these quantities a researcher has to rely on the same econometric techniques that experiments were originally employed to avoid, which knocks experiments off their pedestal and brings them back to reality.

Excited by the experimental turn in social science, the statistician Andrew Gelman elaborates the many ways that experimental reasoning has been and can be incorporated into social research (chapter 7). Though they can never fully save us from having to use techniques of observational data analysis, experiments, Gelman claims, are the gold standard for drawing causal inferences. Nevertheless, he worries that the methodology has become synonymous with randomized experiments and argues that other experimental methodologies can be useful for drawing social science inferences. Gelman thus sits at the center of the controversy over whether, moving forward, all forms of nonexperimental inquiry should be abandoned. This question is at the heart of the book, for if the answer is yes, it is hard to draw any other implication than that everything we have learned in the past is wrong.

Kosuke Imai, Gary King, and Elizabeth Stuart try to rescue both experimentalists and observationalists from themselves by helping to explain and troubleshoot common mistakes made by both (chapter 8). The authors highlight the precise advantages and disadvantages of several observational and experimental research designs.⁸ By considering different techniques such as “matching,” “blocking,” and “randomization” the authors show that different research designs can be employed to deal with different sets of problems. They conclude that both experiments and observational work place constraints on what can be known with certainty.

In chapter 9, Ian Shapiro returns to the larger issues at stake in this book: what promises field experiments can deliver on, and where they fall short. He critiques Gerber, Green, and Kaplan's "research allocation theorem," an implication of which is that resources spent on observational research are more likely to be wasted than resources spent on field experimental research, by arguing that the theorem ignores the possibility of diminishing returns to investments in certain types of research methodologies. Experiments might have produced hard facts that change our intuitions about, say, the effect of campaign phone calls versus face-to-face contact on voter turnout, but the rate at which these discoveries are accruing far outpaces their usefulness to the broader goals of social science. In essence, getting more precise answers to the same questions has opportunity costs in terms of foregone research on other topics.

Shapiro notes that while methods can be used to answer questions, they cannot tell us which questions are worth answering. In letting small empirical findings dictate the next set of questions, rather than letting theoretical questions dictate research programs, the agenda of GGK runs the risk of placing social science on a safe path to nowhere. In the end, Shapiro, by stressing the primacy of a good question, advocates a pragmatic approach to methodological choices.

Read together, the chapters in this book offer a fuller account of the uses and abuses of experiments in social science than one would get from simply reading the experimentalists' tracts. A central theme that unites most of the critics is that the choice of methodology depends in large part on what we want to know. When a large body of theoretical work already exists, experiments may be a good method to test competing hypotheses. But experiments may not be best when staking out new terrain, examining politically or personally sensitive issues, or laying out answers to big questions. My hope is that by arraying a diverse set of views in a single forum, readers, too, will be able to decide for themselves.

NOTES

1. Mill (1843: 440)
2. Mill attributes the concept of varying the circumstances to Bacon, without citation.
3. See Rosenbaum (2002) for an explicit treatment of observational analysis.
4. Researchers in lab sciences and clinical medicine believe that experiments offer advantages over observational techniques, but even here there are some challenges. If

a cancer drug is developed and tested on populations of mice that have been bred to have cancer, we might worry that the results, even if strong for the mouse population, will not apply to humans, bringing into question the external validity of the results. If the same drug is tested on human beings with cancer, some of whom are assigned to receive the drug and others that are not, there is still the possibility for human error (if some people don't take the drug as assigned) or for human intrigue (if some people seek out other treatments at the same time). These issues raise the possibility that the experiment is not internally valid, meaning that the experiment was not carried out exactly as it should be for the results to be reliable.

5. See Henrich et al. (2001).

6. The movement of experiments into the mainstream of social science is evidenced by the increasing publication rates of social scientific laboratory experiments (McDermott 2002, Morton and Williams 2008), field experiments in political economy (Palfrey 2009), experiments in the political economy of development (Humphreys and Weinstein 2009), areas of political behavior and collective action (de Rooji et al. 2009), and development economics (Banerjee and Duflo, this book). Behavioral economics, which since the 1970s has relied extensively on laboratory experiments to test the behavioral foundations of neoclassical economics, has also made movements toward laboratory experiments conducted on location. For a review of this approach, see Camerer et al. (2004). Specific examples of this approach can be found in Henrich et al. (2001) and Habyarimana et al. (2009).

7. Two recent textbooks that espouse the experimentalist view are Druckman et al. (2011) and Morton and Williams (2010). For a classic account see Cook and Campbell (1979).

8. *Research design* is used in a technical sense here to mean a strategy for evaluating data that will produce *causal* estimates of the parameters of interest. See Dunning (2008) for an interesting overview of several approaches to design-based inference.

5

REFLECTIONS ON THE ETHICS OF FIELD EXPERIMENTS

Dawn Langan Teele

1. Introduction

Throughout this book we have considered whether field experiments offer the best way forward for social scientific research. This question has been asked from multiple methodological and epistemological viewpoints, and the chapters reveal considerable disagreement among distinguished social scientists. One issue that has not been raised systematically concerns the ethical implications of running experiments in the spaces where real people live out their lives. By definition, these environments are not wholly controlled by the researcher. Because we, as citizens, might be wary of scientists who, clipboards in hand, descended upon our neighborhoods in order to set up new institutions, start traditions, or change discourses, as social scientists we should be aware of the real-life impacts of field experimental research.

In this chapter I ask whether and how research participants and their communities can be treated ethically in the course of social scientific field experiments. I first establish that by their very natures, experimental and observational studies beget distinctive relationships between study participants and researchers. Both methods raise ethical challenges, but I argue that the ethical hurdles for experimental work are higher than those for observational work. This is because experiments alter the setting, scenery, and sometimes life chances of individuals who are either directly or indirectly involved. Well-known scandals brought these concerns to the fore

in medical research, leading to the development of thorough guidelines for medical research. So far, however, social scientists have assumed that good intentions and an Institutional Review Board's seal of approval suffice to ensure that the dignity and autonomy of research participants are respected.

A critic might respond, even at this early point in the chapter, that the ethics of field experimentation are beside the point, since there is no alternative to field experiments if (1) experimental research is the only way to ensure valid causal inference and rigorous policy evaluations; and (2) the policies being evaluated experimentally would be implemented anyway, but without any sort of evaluation or, worse, with mere observational evaluation. This book contains arguments sufficiently compelling to cast doubt on the first proposition; indeed, the jury is still out as to whether experiments are the only way or the best way to tell us all we need to know about a policy intervention. As to the second point, while it is true that policy interventions should be evaluated so that donors and governments can know whether their resources are being spent effectively, it is misleading to claim that this argument provides an ethical defense of field experimentation as it is practiced today. Not only do researchers often engage in independent experimental projects, without any connection to governments or NGOs, but also they are increasingly engaged in bringing policy ideas to governments and think tanks to be tried out in real-world settings. I don't want to criticize these practices; I think scholars should engage in independent research unconnected to policy evaluation, and that social scientists should play an active role in policymaking, but this means that the ethical implications of field experiments cannot be dismissed by saying that all we are doing is evaluating policies that are going to be implemented anyway.

More generally, although though this chapter raises some ethical issues surrounding the treatment of subjects in field experimental research, it in no way endorses a ban on their use. In what follows we will not encounter any "smoking gun" experiments that cast the method in inexorable doubt. This is attributable to the commendable efforts of Institutional Review Boards in protecting human subjects and to the responsible choices of journal editors in shelving submissions based on ethically questionable

research. But the absence of a smoking gun should be celebrated cautiously, because, as will be shown below, in the writings of the preeminent experimenters a plea can be heard for the prerogatives of research to be placed before the rights of research participants, with some experimenters going as far as to endorse covert experimentation or the deliberate misinforming of human subjects as critical for clean identification strategies.¹ This is particularly troubling given the noble inspiration of many research agendas that employ the experimental method. Research that seeks to learn how citizens can be encouraged to express themselves politically, or that tries to understand how villagers can pull themselves out of poverty, is fundamentally and sincerely concerned with improving lives and welfare. If the method used to address these questions involves misleading potential participants and ignoring the sometimes diffuse risks that experiments pose to communities in a research site, it undermines the very ideals that originally inspired its pursuit. For this reason it is imperative that the question of experimental ethics be brought into what is all too often an entirely technical methodological discourse.

The chapter is organized as follows: section 2 argues that the entry of researchers into real-world settings and the manipulation of subjects in these environments, which distinguishes field experiments from other research methodologies, raise specific ethical dilemmas and increase the ethical burden of justification for field experimental researchers. Section 3 describes and applies the principles of the *Belmont Report*, developed to guide medical research, to field experimentation in the social sciences. It demonstrates that the spirit of these principles has been violated in the writings of some prominent experimenters. In conclusion, section 4 discusses prescriptions and suggestions for how field experimental practices could be made to comply with ethical principles. I suggest that informed consent and a more thoroughgoing evaluation of the downstream and community-level risks that stem from field experiments must guide all research if it is to be ethical. Where scholars are worried that informed consent will change the behavior of participants and thereby interfere with the estimators of causal relationships, I suggest that placebo group designs will help the research remain inside the bounds of acceptable treatment of human subjects.

2. What Makes the Field Experimental Method Different from Any Other?

How do field experiments differ from other research methodologies that involve studying people—cataloging their responses, measuring their movements, and interpreting these data? The most basic difference is that whereas observational research hopes to make causal inferences by measuring and analyzing variation in the world, field experiments induce the variation whose outcomes will later be studied. In a classic article on causal inference, Holland (1986: 959) famously writes, “No causation without manipulation,” implying that a researcher cannot claim that a causal relationship exists between a catalyst and an outcome unless the catalyst is set into motion by an experiment. Though the use of the term *manipulation* seems rather anodyne, manipulation of the research environment is precisely what makes field experiments different from, and potentially more ethically dubious than, observational work. To illustrate this point: if observational social scientists are spectators of a card game whose hands nature dealt, experimental social scientists have positioned themselves as the dealer. This shift from spectator to dealer changes the relationship of the social scientists to the players, and it begs for an examination of the practices and policies that bind them together.

Table 1 makes this distinction clearer; it catalogues two dimensions of social science research: *Field* (on the vertical axis) and *Intervention* (on the horizontal axis). Elizabeth Wood (2007) defines field research as research that is “based on personal interaction with research subjects in their own setting” (123). Thus *Field* in this table signifies whether, in order to carry out a project, a researcher must join the “setting” of the research participants. Research that does not involve joining someone else’s setting includes the use of data that can be downloaded from official websites, the collection of records from archives, and research in a laboratory environment.

Intervention, the variable along the horizontal axis, signifies whether the researcher purposively manipulates the research context in some way, for example, by randomly assigning participants to treatment and control groups. A heuristic for understanding this dimension is the answer to the question, Does the research itself alter the environment in which the

Table 5–1. Classifying Social Science Research by Intervention and Entry

	Intervention: No	Intervention: Yes
Field: No	Archival or library work	Lab experiments
Field: Yes	Field surveys, ethnographic work	Field experiments

participants are active? Or, perhaps more accurately, Does the research *intentionally* alter the environment in which the participants are active? The answer to these questions is clearly no for work with national accounts data and archival research and for surveys carried out in the field (even surveys that randomize the order of questions). The answer is also no for ethnographic work, in which a cardinal rule is for the researcher to avoid actively influencing events.² To sum up, the answer to the question of whether intentional alteration occurs, in other words, manipulation of the research environment, is yes for social science experiments carried out in the lab and also for those carried out in the field.

The point I wish to highlight from table 1 is that moves down and to the right increase the burden on the researcher to question whether the research comports with ethical principles. Research that moves down the vertical axis of *Field* must be subject to more critical scrutiny because interacting with research participants in their own setting can exacerbate existing power dynamics because of the often large cultural, educational, and socioeconomic differences between the researcher and the participant. Ethnographers in particular are sensitive to these power dynamics. Standard practice in research that employs the tool of participant observation mandates that certain groups must be studied with care: the participation of those who are economically vulnerable, mentally handicapped, socially deviant (Thorne 1980) or who live under authoritarian regimes (Goduka 1990) has been given considerable attention by ethnographic methodologists. Wood (2006) summarizes a basic message from this literature as doing no harm, meaning that the researcher has to anticipate and be attentive to the ways in which the research itself might complicate the lives and practices of the people being studied. To put it another way, the consensus among those who employ the techniques of participant observer conceive of ethical field research as similar to deep woods camping: the research itself should leave no trace.

Research that moves across the horizontal axis of *Intervention* also requires greater ethical scrutiny because, by definition, it undermines the leave-no-trace maxim. Indeed, experimental intervention intentionally alters the participant's environment and can produce unintended and unforeseen consequences. Unforeseen consequences have typically been examined in the context of deception: the famous psychological experiments by Stanley Milgram (1974) first inspired inquiries into whether deception is an ethical practice in social research, and later works by Baumrind (1985), Geller (1992), and Bonetti (1998) have followed up on the negative outcomes from experimental deception.

Milgram was interested in understanding how seemingly decent people could, under the sway of an authority figure, be persuaded to inflict pain upon another person.³ He devised an experiment wherein one research participant acted as a "teacher" who was to help another participant, the "student," learn a list of words. If the student answered the teacher's prompt incorrectly, the teacher was to administer an electric shock to the student with increasingly higher voltage as the total number of wrong answers increased. Two things were key in this experiment; first, the teacher could hear but not see the student, who was strapped to a chair, with electrodes attached to his wrists, in a separate room. Second, a "medical doctor," who served as the experiment's authority figure, was present in the same room as the teacher. The doctor encouraged the teacher to keep administering electric shocks to the student when the latter produced the wrong answer. Milgram's experiment revealed that most of the teachers, despite voicing concerns and hesitating to administer electric shocks, continued to issue the shocks when the doctor encouraged them to do so. Many even continued to shock the student until the maximum voltage was reached.

The reaction to the Milgram experiment was visceral. Many questioned the ethical validity of the experiment on the grounds that the deception involved in the experimental protocol inflicted an unusual amount of stress on the research participants (Baumrind 1985). Alternatively, Patten (1977) argues that Milgram's experiment was unethical because the researchers claim that the stress to the participants was *unintended*, when, insofar as the participants had to believe they were inflicting real pain for inferences to be valid, the very nature of the experiment depended on this discomfort.⁴

Two issues spring from the Milgram debate that are relevant to our purposes here. The first is that experiments can inflict undesirable consequences on participants, in this case stress and anguish. The second issue is that some of these consequences will be unforeseen by the researcher. It is precisely these negative consequences, both foreseen and unforeseen, that increase the burden on the experimenter to scrutinize the ethics of the intervention.

What differentiates the field experimental method from any other, and which is revealed by its location in the bottom right quadrant of table 1, then, is that the method raises ethical concerns both by entering into a research participant's setting and by intentionally manipulating the research environment. The ethical terrain is therefore complicated by the very nature of the method.

Despite this, even as field experimentation became more common in social science, there has not been much discussion of its ethical requirements. Most papers promoting the method contain some statement that experiments are the ideal method, provided they are ethical (Banerjee and Duflo, chapter 4 in this book; Humphreys and Weinstein 2009), but only few actually describe the nature of this responsibility (for an exception, see Barrett and Carter, chapter 3 in this book). Further, while there are well-developed standards for reporting measurement strategies and estimation protocols, there are no current standards for reporting on the ethical challenges involved in the research process.

To get some traction on these issues, I provide below a nuanced reading of a foundational document in the ethics of human subjects' research, the *Belmont Report*, and pinpoint places in which the report's three principles—respect for *human dignity*, *beneficence*, and *justice*—have been undermined in the new experimental social science. The *Belmont Report* is a foundational document because it is cited in the bibliographies of most guidebooks for institutional review of human subjects' research. A cursory survey of the institutional review documents at fifteen top research universities—including, inter alia, Yale, Harvard, Cornell, Stanford, and Northwestern—reveals that every one of these universities lists the *Belmont Report* in either the mission statement or the opening page of their human subjects' manuals. Thus, the report is a useful launching pad for my discussion of experimental ethics.

3. Implications of the *Belmont Report* for Field Experiments

Experiments have been a tool of medical research dating back to the second century A.D., and many medical innovations, including the development of the invaluable smallpox vaccination, might not have been possible without some form of experimental science (Ivy 1948). Nevertheless, unpalatable choices have been made in the course of experimental science, causing the professions that utilize this technique to exert a considerable and commendable effort to outline and enforce the ethical and legal limits of this type of research.

An infamously unpalatable example that led directly to the construction of ethical codes for human subjects' research is the Tuskegee Syphilis Study, which was carried out by the U.S. Public Health Service from 1932 to 1970. The experimental population was composed of 399 black men, 199 of whom had syphilis, from very poor rural areas near Tuskegee, Alabama. In exchange for routine checkups and regular blood work, participants in the experiment were given access to clinical care unrelated to the disease, were fed warm meals on days that they went to the lab, and were promised a stipend to defray the costs of burial in the event of death. The purpose of the research was to study untreated syphilis as it progresses, and all members of the treatment group were in late stages of the disease.

There are several ethical problems embedded in the Tuskegee study. To begin with, participants were not told they were being called on to study syphilis, a disease which many of them did not know they had. Rather, participants were told that they had "bad blood," a colloquial term for syphilis, that the doctors wanted to test, and some participants actually believed they were being treated for bad blood rather than contributing to the study of syphilis (Jones 1981). Despite the emergence of penicillin (a reliable cure for the disease) in the 1940s, none of the study's participants were actually administered treatment for syphilis. Years of investigation revealed that the scientists in the Tuskegee study deliberately withheld information about the disease and possible treatments in order to study its effects as it progressed (ibid.).

As documented by Jones, when the American public was informed about the Tuskegee study in 1972, the overwhelming response was outrage. The U.S. government, which should actively protect its vulnerable citizens, had instead inflicted a considerable amount of pain and harm on them. Moreover, it could hardly escape notice that the study's subjects were members of a marginalized racial minority that was largely illiterate and extremely poor. This aroused suspicions that the study was carried out on those who would not know that their rights were being trod upon and who were likely to feel grateful for whatever remuneration they were offered. In other words, the researchers in the Tuskegee study chose their study population precisely because, due to their marginalized social and economic status, subjects were unlikely to learn about the nature and treatment of their disease and, in the event they did, were unlikely to take action against the researchers. Jones writes, "The ultimate lesson that many Americans saw in the Tuskegee Study was the need to protect society from scientific pursuits that ignored human values" (1981: 14). Though not the first and clearly not the last time that the rights of minority citizens would be violated in America, the Tuskegee study raises the specter of discrimination in the choice of study populations that remains as worrisome today as it was in the 1930s.⁵

To overcome the apparent ethical blind spots of the scientific and medical communities, in the late 1970s the U.S. Department of Health, Education, and Welfare formed a National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. The commission, which was made up of lawyers, legislators, academic researchers, and physicians, convened to establish guidelines for research on human subjects. The fruit of this conference, *The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research* (1978), is meant to serve as an "analytical framework that will guide the resolution of ethical problems arising from research involving human subjects." The *Belmont Report* is a foundational document in the history of institutional review practices: it does not make technical recommendations for how to evaluate the ethics of particular research projects, but rather enumerates three principles to guide ethical research with human subjects: respect for persons, beneficence, and justice. In

the following pages I will demonstrate that the concrete examples in the *Belmont Report*, which takes biomedical and behavioral researchers as its audience, are not exhaustive of the challenges that arise with regard to field experiments. My aim is to interpret the principles of the report in light of experimental research, so that social scientists, in so far as they endorse these principles, will have an ethical framework on which to base experimental research.

3.1. *Respect for Persons*

According to the *Belmont Report*, respect for persons entails acknowledging that the subjects of experiments are autonomous individuals whose personal dignity and opinions must be considered and respected. This sometimes means protecting those whose autonomy is in some way diminished—by age, mental handicap, or other debilitating condition—and it generally requires that a participant's anonymity be protected. Moreover, a researcher must not obstruct a participant's ability to make a well-reasoned judgment about whether to participate in a given study, implying that she cannot withhold any information that may be relevant for the participant to give informed consent. In certain situations, including those where a participant is vulnerable, immature, or incapacitated, the report requires a researcher to take extra steps to ensure that the subject is protected.

Adhering to this principle is relatively straightforward: researchers must communicate the nature and level of risk associated with participation in the research and must procure a statement of consent that demonstrates the subject's awareness of the potential risks associated with participation and that confirms that the subject participates willingly. In all research, application of the respect for persons principle requires that the researcher have some idea of the risks associated with the treatment. I will speak more about the issue of risk in the next section, but for now suppose that the researcher is aware of the risks. For example, in biomedical trials a patient must be informed of potential side effects of a drug (that it could cause nausea, headaches, fever, and so on) and in behavioral studies the researcher must communicate essential facts about the research design—for example, that the subject may be exposed to loud noises, confined spaces, or some painful stimuli—before obtaining consent.

It is relatively clear how respect for persons translates into informed consent in biomedical and behavioral research, but in the case of field experiments its application is not straightforward. Many scholars, including Banerjee and Duflo (chapter 4 in this book) and Levitt and List (2009), point out that if participants are aware that they are participating in a field experiment, they may act differently and in such a way as to confound measurements and results that stem from the project. This phenomenon, which applies to all situations in which human behavior is measured, is widely known as the Hawthorne effect.⁶ Though the problem is not unique to field experiments, it is particularly vexing to field experimenters, whose claim to methodological primacy rests on two promises: first, that there is internal validity to the research design, and, second, that opportunities for extrapolation to other contexts—external validity—are less compromised with field experiments than with other methods. The Hawthorne effect poses problems on both levels, as subjects' responses are due to other influences besides just the treatment, undermining internal validity. Second, if a subject acts differently when he knows he is being studied, then the study's findings will likely not apply to other nonexperimental contexts. Faced with these obstacles, many field experimenters argue for the use of "covert experimentation," in which people are unaware that they are participating in an experiment and are sometimes unaware that their behavior is being measured at all (Levitt and List 2009).

Here we are presented with a problem: if it is true that informed consent compromises both internal and external validity and thereby undermines the scientific integrity of social research, field experiments that necessitate forgoing procedures of informed consent fail to satisfy the first principle of respect for persons.⁷ Humphreys and Weinstein (2009) speak of this dilemma as a trade-off between ethics and measurement, an ethical dilemma without a clear-cut solution. But this misstates the problem. An ethical dilemma arises when it is impossible to simultaneously meet the demands of two *ethical* principles, as, for example, when one is confronted with a situation in which lying to a friend is the only way to avoid insulting him. The ethical principles that conflict in this example are not to lie and also to be kind to others. The trade-off created by the Hawthorne effect between satisfying the principle of respect for persons by obtaining informed consent and generating unbiased measurements of

causal effects does not have this character. It is more akin to the trade-off, in criminal justice, between respecting a suspect's Miranda rights and the prospects of securing evidence that will lead to conviction. In the latter trade-off, one can recognize that there may be some conceivable benefit to denying suspects their rights, while still demanding that the principles embodied in the Miranda rights be satisfied; the problem here is not an ethical dilemma, but one of comporting with an ethical principle or not.

Given that the most important implication of the principle of respect for persons is that researchers gain informed consent from their subjects, field experiments that do not satisfy this demand fail to adhere to the principle and, as a result, are unethical under the principles of the *Belmont Report*. In the American criminal justice system, evidence gained through the use of improper procedures (for example, entrapment, denial of a lawyer, etc.) is inadmissible in court. One might argue that, by analogy, evidence procured in field experiments for which informed consent of subjects was not secured should not be admissible in peer-reviewed journals, the courts of scientific discussion.

There are two compelling counterarguments to the strong claim made above that all participants in field experiments must give informed consent. First, some researchers argue that given the everyday situations that are the site of behavioral studies, requiring informed consent would be not just impossible but in fact ridiculous.⁸ What would life be like if supermarkets were forced to gain informed consent before they analyzed data on price changes and purchasing behavior? How could banks operate if they were forced to gain consent before they investigated responses to marketing letters? The tenor of these objections is that there are numerous situations in which behavior is already being manipulated and monitored and that to subject academic researchers to a standard different from that of private businesses is unfair. This objection has been confronted in work on ethical ethnography (Bosk and Vries 2004) and in the medical sciences (Gray 1978), where a similar resentment of academic research's comparatively stringent ethical principles has been expressed.⁹

But this logic should be rejected for two reasons: first, there is something fundamentally different about supermarket chains using experiments on nonconsenting shoppers to determine whether they are more likely to buy a product priced at \$0.99 than \$1.05 and a social scientist

approaching an NGO or a village with an experiment that she seeks to implement. Upon entering a store, the shopper submits himself to the rules of the store, but the same is not necessarily true when an experimenter manipulates incentives in a real-world context (especially when she experiments on people outside of her own community, which is generally the case). Second, and more important, many research agendas are undergirded by a desire to promote positive changes in the world—higher levels of schooling for girls, less discrimination in the political sphere, human rights for the world’s poorest—a motivation that would be compromised if the standards of the *Belmont Report* were neglected in favor of the standards of the business world.

In a second, related objection, some researchers contend that informed consent is unnecessary when the risks of a particular experiment are negligible. For example, it could be argued that there is absolutely no risk involved in nonpartisan get-out-the-vote experiments in which some people receive automated telephone calls reminding them to vote (the treatment) or to recycle (the control) (see the experiment in Gerber et al. 2009). In these risk-free scenarios experimenters argue that requiring informed consent would be an onerous bureaucratic hardship that hinders the advance of social scientific knowledge for no particular reason. But calculations of risk are independent of respect for individual autonomy, and the principle of respect for persons demands that people’s judgment be respected regardless of the potential costs or benefits that the research exacts upon them. In other words, respecting a person’s autonomy requires that participation in research be voluntary, and so, independent of the costs or benefits of the research, the researcher does not get to choose for the volunteer how much risk is tolerable.

Moreover, it should be noted that many people choose not to volunteer when asked to participate in a randomized experiment: were they asked for consent, some people would say no. Levitt and List write, “It is commonly known in the field of clinical drug trials that persuading patients to participate in randomized studies is much harder than persuading them to participate in non-randomized studies. . . . The same problem applies to social experiments, as evidenced by the difficulties that can be encountered when recruiting decentralized bureaucracies to administer the random treatment” (6). The authors identify this as the problem of

randomization bias—in which the experimental pool is not representative of the population because some types choose not to participate. They advocate covert experimentation as a way to circumvent this type of bias. Advocating for covert experimentation does not seem like a tenable long-term solution to the problem of unwilling volunteers. Precisely because randomization bias indicates that people do in fact have preferences over their involvement in experimental social science, this means that no matter how great the potential gains to the research community, ethical research cannot put the prerogatives of the researcher over a person's right to volunteer or to decline to participate.

3.2. *Beneficence*

The second principle outlined in the *Belmont Report* is the principle of beneficence. Beneficence is an obligation that researchers have to secure the well-being of their subjects, which includes the obligation not to expose subjects to “more than minimal risk without immediate prospect of direct benefit” (*Belmont Report*: 6). In particular, the researcher must consider the nature and degree of risk to which the subject is exposed, the condition of the population that will be involved in the research, and the level of anticipated benefits for that population. In biomedical research, extensive pretesting (for example, on animals) is performed before a drug is approved for human testing,¹⁰ and thus a researcher should have a good idea of what types of risk a given subject will face.

But the concept of risk may be more difficult to assess in social science field experiments than in biomedical or behavioral research. One reason is that social experiments take place in real-world settings, that is, within thick social, economic, and political contexts, the nuances of which often elude even a well-informed researcher. To give a concrete example, it is often argued that women who do not earn wages do not have bargaining power in the household equal to their wage-earning husbands, fathers and brothers. Targeted microcredit—small-principal group lending that is marketed to women in the developing world—has long been hailed as offering a path to women's empowerment by giving women access to capital of their own (e.g., Karlan et al. 2006).¹¹ At first glance, then, microcredit programs and the numerous field experiments structured around them seem unproblematically beneficent: women are subordi-

nated because they have no capital; give them capital and right this social wrong.

But long-standing entitlements, patriarchy included, do not die swiftly. Programs that undermine age-old hierarchies in the family may provoke the violent behavior the program hoped to avoid (Schuler et al. 1998). Indeed, anthropological work from Bangladesh shows that, contrary to targeted microcredit programs' intentions, husbands are often the primary users of loans. More worrisome still, women who participate in microlending programs face increased risks of personal victimization after gaining this new financial access (Rahman 1999). The point to emphasize here is that while the risks of certain experimental and development initiatives might appear minimal, experiments conducted in complex social contexts involve risks that are unpredictable and even unknowable *ex ante*.¹²

Applying this logic to another context, consider a recent article by Paluck and Green (2009). The authors assert that social atrocities like the Rwandan genocide are facilitated by a political culture in which dissent is frowned upon and where authority is blindly followed. They argue that Rwandan society would be better off if more dissenting opinions were held and voiced. The authors carry out a field experiment in Rwanda to test whether radio soap operas that feature dynamic plots and dissenting voices inject this language into the discourse of those who listen to the program. The experimental results indicate that the group that received the dissent-imbued treatment was in fact more likely to voice opposing opinions and less likely to turn to authority figures to solve conflicts. The paper, which certainly strives to promote the widely held democratic value of contestation within Rwandan society, fails to consider the possibility that the experiment might carry some physical risk to the people who participate or invite violence in the community as the structure of its social fabric is rewoven.

But if extending microcredit to women can increase violence against them, it is not inconceivable that introducing the language of dissent in a postgenocidal society might carry similar risks. I should note that, to their distinct credit, Paluck and Green take care in appendix I (638–39) to discuss the “Procedural and Ethical” details of the experiment.¹³ However, their discussion does not ask whether introducing the language of dissent

brings any new risks to the community; their concerns were limited to the (important) issue of whether participants might experience psychological trauma when they speak of the genocide.¹⁴ The issue to highlight is that no matter how well intentioned the researchers, which Paluck and Green certainly are, there are many contingencies in social environments. Understanding the risks involved in an intervention, in line with the *Belmont Report's* principle of beneficence, is therefore not an easy task.

Besides unforeseen risks, a second concern raised by the principle of beneficence in a field experimental context is the uneven allocation of goods both within and between communities. In recent reviews of the emerging field experiments literature, both Banerjee and Duflo (chapter 4 in this book) and Humphreys and Weinstein (2009) acknowledge that many experiments necessitate doling out goodies to some and not to others. Both papers note that in many situations the researcher must confront and deal with bad feelings or jealousy among participants. According to Humphreys and Weinstein, jealousy is problematic because “differences in outcomes may be interpreted as evidence for a positive program effect on the treated community even if all that has occurred is an adverse effect on the control community” (2009: 375). In other words, jealousy is acknowledged as a potential problem for field experiments because it may alter behavior in the control group and confound precise measurement. But this rationale ignores the reasons that jealousy arises in the first place as well as its ethical implications: a randomized social intervention benefits some villages or villagers and not others. Even if this benefit is small, it is difficult to know ahead of time how it will be perceived or how long the changes in behavior it occasions will persist.

Banerjee and Duflo address this issue with the following advice: “Implementers may find that the easiest way to present [the project] to the community is to say that an expansion of the program is planned for the control areas in the future (especially when such is indeed the case, as in phased-in design)” (chapter 4 in this book, p. 000). Note first that the authors’ suggestion that control villages should be told that the project will eventually be extended to them “especially” when it is true implies that in some cases it might make sense to mislead the villagers, suggesting that they will eventually receive some treatment even when the researchers know the program will never be extended. Banerjee and Duflo’s

work has been some of the most important in development economics over the last twenty years—their is both pathbreaking and compassionate scholarship that has affected development initiatives and undoubtedly done much good in the world. But their statement in the above quotation, besides being problematic in its own right, reveals a deeper issue: even when researchers are aware of the many levels on which social conflict can be introduced by experimental interventions, they treat it as a methodological rather than an ethical difficulty.

Moving forward, it is important that concerns about what is ethical not be solely interpreted as methodological problems. For this to happen, researchers will have to actually wrestle with the principle of beneficence. They need a fuller sense of the context and a more thoroughgoing commitment to understanding the risks of their research in these contexts. Banerjee and Duflo (chapter 4 in this book) claim that while scientifically interesting, measuring equilibrium effects can be extremely difficult. The principle of beneficence requires that we at least try.

Contrasted with biomedical and behavioral research, where community-wide equilibrium effects may be extremely rare, it is clear that field experiments pose particular challenges for understanding and communicating risk, especially when people who are not chosen as participants may be affected. If this is true, there can be little confidence that community-wide risks can be adequately accounted for in the design of field experiments. Perhaps by teaming up with ethnographers and social workers, field experimenters will be able to command a more thorough understanding of possible downstream effects of their interventions, that is, of the effects that an intervention might have on social processes aside from those that are part of the study.¹⁵

3.3. *Justice*

The last principle elaborated in the *Belmont Report* concerns justice: “Who ought to receive the benefits of research and who ought to bear its burdens?” The report acknowledges that there are competing ethical principles of resource distribution that could reasonably be called on when considering the benefits of research, but it emphasizes the following issues: first, the benefits of research should not accrue disproportionately to a class or race that does not itself participate in the research.¹⁶

And, second, “the selection of research subjects needs to be scrutinized in order to determine whether some classes (e.g., welfare patients, particular racial and ethnic minorities, or persons confined to institutions) are being systematically selected simply because of their easy availability, their compromised position, or their manipulability, rather than for reasons directly related to the problem being studied” (*Belmont Report*: 7).

As to which populations will benefit from field experimental research, there is good reason to believe that most experimental initiatives in developing countries are intended both to understand and to alleviate poverty. Indeed, results-based development initiatives are concerned primarily with how to allocate the tremendous amount of money that goes to international aid in a socially desirable and economically efficient way. Hence, many experiments are designed to evaluate the efficacy of a program for the people who are beneficiaries of the program. Some of this research has uncovered evidence that certain development practices are inefficient. For example, Gugerty and Kremer (2008) find that development aid that is channeled to women’s organizations in Kenya has the undesirable consequence of pushing less-educated, poorer women out of leadership positions. The authors argue that what seems good in theory (more money to community organizations) might be harmful in practice. In biomedical trials justice often entails ceasing the trial immediately if a drug is found to be harmful to participants, or making the drug available to all participants if the beneficial effects were found to be overwhelmingly clear. By analogy, the Gugerty and Kremer findings might, under the principle of justice, require that the funding be redirected toward areas that better serve the purpose of the initiative.

But though much of the field experimental work in the developing world has hitherto been associated with program evaluation, in other words, with measurement of treatment effects of public policy initiatives, most academic researchers would like to shift the trajectory toward initiatives that they propose themselves (Banerjee and Duflo, chapter 4 in this book; Humphreys and Weinstein 2009). As both Deaton (2009) and Heckman and Smith (1995) argue, academics cannot be concerned solely with “what” works from a policy standpoint, but also with “why” it works. They claim that academic research is not meant only to find out how to make some desired end come about through some available

means, the goal of public policy evaluations, but rather to uncover the social and behavioral characteristics that undergird human interaction.

To the extent that field experimental research becomes detached from policy evaluation, complying with the principle of justice will require strong defenses that hypotheses being tested are specific to the population under study. Selection of subjects from either economically vulnerable groups (who may be more willing to participate because of the relative value of money or goods-in-kind offered by researchers) or politically vulnerable groups (who are less likely to challenge the authority and motives of a research scientist) will demand heightened scrutiny.¹⁷ These concerns are particularly relevant for field experiments in developing countries, where subject populations are often both politically and economically vulnerable.

This is not an idle worry. In a frank admission, Banerjee and Duflo note that the particular conditions of people in developing countries make them good subjects for experimental research: “Limited government budgets and diverse actions by many small NGOs mean that villages or schools in most developing countries are used to the fact that some areas receive certain programs whereas others do not. . . . *When the control areas are given the explanation that the program has enough budget for a certain number of schools only, they typically agree that a lottery is a fair way to allocate those limited resources. They are often used to such arbitrariness and so randomization appears both transparent and legitimate*” (Banerjee and Duflo, chapter 4 in this book, 000; emphasis added).

The authors assert that because people in developing countries are used to “arbitrariness” in decision making, they will likely accept the randomization protocol without too much fuss. Said somewhat differently, the fact that life is unfair for the poor in the third world means subjects might not only agree to participate in a randomized experiment, but also will believe that randomization is a “fair” way to allocate resources.¹⁸ Moreover, as a telling footnote reminds us, “It should also be noted that the lower cost of the programs and working with NGO partners greatly expand the feasible set of experiments in development, compared with what has been feasible in the United States” (Banerjee and Duflo, chapter 4 in this book, 000). By these lights, experimentation in the tropics is pursued not only because these populations tend not to be hostile toward researchers,

but also because the costs of studying them are lower.¹⁹ These sentiments clearly violate the spirit of the *Belmont Report*. To put it bluntly, taking advantage of populations' contextually induced vulnerability to random interventions and the reduced expenses that result from their relative poverty is simply and obviously unethical.

There are many instances in which a research question absolutely must be answered outside of the researcher's own community. For example, in a series of laboratory-like experiments in the field in fifteen small-scale communities around the world (hunter-gatherer groups, tribes, and so on), Henrich et al. (2001) study whether the assumptions of the neoclassical self-interested actor hold up in different social and economic contexts. A large body of research based on experiments with undergraduate students in the United States and around the world shows that the neo-classical self-interested actor may be the exception rather than the rule. In their study Henrich et al. (2001) hypothesize that these results will hold up across cultures and seek to test this in many different social contexts. Their research question thus drives the selection of their study populations, not the other way around. It is important to ask, then, in light of the justice principle elaborated in the *Belmont Report*, why certain projects must be carried out on populations far away from the researcher's home. In the very cities of American universities there are low take-up rates of social services, collective action failures of trash pickup, low savings, and other phenomena that might be of interest to scholars. A good justification, not just one of convenience or expense, should therefore be given for why a particular research question cannot be answered other than by experimenting on the economically and politically vulnerable populations of the developing world.²⁰

Finally, the principle of justice serves to remind researchers that they are often in a powerful position relative to those they study. When an experimental intervention has some monetary or in-kind benefit associated with the treatment, social scientists should recall that uneven access to resources and the ability to decide who gets what and when are fundamental privileges of power. When social scientists enter into research relationships with governmental officials, NGO employees, and villagers the world over, they must consider the nature of their power in these contexts. Their power, which stems from their role as liaisons between

cultures, also exists because research results influence international aid, public health initiatives, and foreign policy in nontrivial ways. This is true not only when academic researchers partner with organizations like USAID, the World Bank, and local governments; it is also true because their research, which is open to consumption by journalists and public intellectuals, can influence the views of the public at large.

A skeptical reader might argue that power differentials arise in all contexts in which a researcher studies vulnerable populations at home or abroad and that it is independent of the experimental research design. This is true: ethnographers who study tribes in the Brazilian rainforest, sociologists who study deviant behavior, and tenured laboratory experimenters working with undergraduate subjects might all be relatively advantaged vis-à-vis their research subjects. But field experiments are set apart because, as I argued in section 2, the object of the experiment is often not to observe people in their day-to-day lives but to intervene in a purposeful manner to affect social change. Moreover, many field experiments, especially those in which a researcher has partnered with an NGO or a development organization, are meant to actually change people's behavior in a direction that the organization and researcher deem more socially or economically desirable. This is a very different model from that followed by a participant observer—it is closer to social engineering than to social studies. Justice should therefore be a primary concern for experimenters who seek ethical interaction with their research populations.

4. Conclusion

Field experiments differ fundamentally from laboratory experiments or participant-observer field research in ways that compromise the ethical integrity of the method. The adage “No causation without manipulation” requires that experimenters alter something in a real-life setting in order to recover unbiased causal estimates of social relationships. The field experimental method also differs from biomedical research because, in the thick social contexts where they are carried out, community-level risks become a real concern. I have understood these differences as a call to reinterpret the principles of the *Belmont Report* with an eye to guiding ethical use of this method of social research. The principles—respect for

human dignity, beneficence, and justice—should be adhered to in good faith by all field experimenters. Importantly, this is not to preserve the integrity of the document but rather because the principles are based on values that most contemporary researchers share.

Where ethical considerations might pose problems for experimenters, for example, if being observed influences behavior (the Hawthorne effect) or if knowledge of the randomization leads to few volunteers (randomization bias), experimenters must be more creative. Measurement in medical trials was compromised when physicians and patients were aware of treatment assignments. This problem is commonly addressed by double blinding, a practice that leaves physicians, nurses, and administrators in the dark about treatment assignments. Patients, for their part, consent to being in the trial, even if they are given a placebo rather than the treatment. To maintain high ethical standards, field experiments could also utilize placebo groups. This would allow for informed consent—volunteers agree to participate in the experiment—though they won't know what group they are in. Note that informed consent here does not require that participants know up-front what is being measured, as people can reasonably commit to not knowing the object of study while agreeing to participate in the study. Nevertheless, this commitment should be countered with commitments by the researcher: to bring the participants in on the study's purpose after the trial is over and to share research results after the report is finished. This common practice in anthropological work would be a welcome addition in the context of many research contexts.

Researchers also need to do a better job of understanding the potential downstream consequences—the community- and individual-level risks—associated with field experiments. This requires knowing a good deal about the people and groups that are being studied and may necessitate teaming with ethnographers or social workers to fully think through the ways in which perceptions of fairness and jealousy that are potentially induced by the experiment might pull at the social fabric in the community under study. Interestingly, where field experiments have been thought to offer a single methodological key to social scientific problems, the ethical challenges of the research may require working with scholars in other fields with other skill sets. It seems that mixed methods are here to stay.

Finally, in terms of justice, high standards must be set for a defense of the chosen subject population. The chosen population must be direct po-

tential recipients of the research; thus, there have to be very clear reasons why the poor or the vulnerable in ones' own community or abroad are the subjects under study. That these groups are cheaper to study—either because money goes a long way in poor countries or because the gatekeepers are more lax about access to their citizens, subjects, or community members—is very far from the mark of ethical best practices.

In closing, I want to stress that the objective of this chapter is not to claim that more stringent regulations should be imposed by Institutional Review Boards across the country. Perhaps they should, but my concern lies elsewhere. I seek, rather, to convince us to be socially conscious in our role as experimenters. We need not think that the principles of the *Belmont Report* are barriers to scientific discoveries. Rather, adherence to its principles allows our values to be borne out within our research agendas. Scholars of development, governance, ethnic politics, and collective action should be particularly attuned to the problems that arise when some voices are given too little weight relative to others. Therefore, insisting on informed consent, full assessment of risk, and nonexploitative participant selection procedures are minimal steps toward ensuring that experimental social science lives up to its noblest aspirations.

Acknowledgments

This essay has benefited enormously from the guidance and encouragement of Elisabeth Wood, Nick Sambanis, Susan Stokes, Donald Green, and Matthew Kocher. I would also like to thank Peter Aronow, Blake Emerson, Navid Hassanpour, Malte Lierl, Rohit Naimpally, Erin Pineda, Joshua Simon, Alison Sovey, and Rory Truex for their thoughtful comments on earlier versions of the chapter. Many of the counterarguments that appear here are the result of conversations with Dean Karlan, Donald Green, and Abhijit Banerjee, and several anonymous reviewers whose strong positions have, I hope, been given adequate audience here.

NOTES

1. I am not talking about the use of deception as a tactic within an experimental framework, whereby participants are misled as part of the experiment itself. The concern is rather with deception that is used before the participant has agreed to participate in the project. More on this below.

2. This leave-no-trace heuristic for ethnographic work is accompanied by a large body of post-fieldwork reflections on ethical issues (Whyte 1988; Barnes 1977: 34).

3. In his book, Milgram describes his inquiry as stemming from Hannah Arendt's work on the trial of Adolph Eichmann (1974: 175–78). His theoretical question asks how a seemingly ordinary person like Eichmann could carry out tasks that he knew led to the extermination of many of Europe's Jewish residents. The answer that Milgram's experiment provides is that ordinary people might do abhorrent things in the face of authority.

4. Patten also contends that the experiment is unethical because participants are encouraged to do something that is "shockingly" immoral (1977: 350), a concern I would also raise in a discussion of recent work in political science, such as Lagunes et al. (2010).

5. Skloot (2010) reports on the tale of Henrietta Lacks, a black woman who died of cervical cancer in 1952. Her "immortal" cancer cells were the first ones scientists were ever able to grow in the lab, a process that contributed, among other things, to the polio vaccine (2010: 188). There are ethical and legal questions about the proper remuneration for the original owners of cells, but for our purposes the disturbing ethical question comes much later, in 1973, when the First International Workshop on Human Gene Mapping, held at Yale, decided to procure blood samples from the surviving children of Henrietta. A postdoctoral fellow carried out orders to get the blood samples from the Lacks children, with no instructions to inform the family of why the samples were needed (182). The unfortunate Lacks children, themselves having less than primary school education, thought they were talking to doctors in order to be tested for the cancer that killed their mothers. It is clear in Skloot's account that the family was deliberately misled and that the scientists that handled this case were not forthcoming with answers about their research and their need for the Lacks's involvement. Needless to say, the Lacks children were not treated with the dignity befitting human beings.

6. The Hawthorne effect was discovered in 1939 during a sociological study of the Western Electric Plant outside of Chicago. Barnes concludes that the discovery "was the first significant nail in the coffin of the natural science paradigm as used in social science, for it drew attention to the fact that the interaction between scientist and citizen is two-way, and that the process of inquiry itself has consequences for both parties" (Barnes 1977: 46).

7. Interestingly, there have been many discussions in the field of psychology as to whether it is ethical to deceive research participants that participate in laboratory experiments. Sometimes deception itself has been the object of inquiry, but often it is used so that participants do not alter their behavior based on a desire to help the researcher. In an interesting review of the deception question, Bonetti (1998) argues that deception should be part of the experimental economists' toolbox, but, more important, he also gives evidence that participants who are *not* deceived do not act much differently than those who are. This relates to my discussion because the experimenter's worry that informed consent will bias behavior may be unfounded. Thus

there is at least some evidence that forgoing informed consent may not be justified on measurement grounds.

8. I am indebted to Dean Karlan for raising this objection in conversation.

9. Gray notes that “one can still find an academic surgeon wistfully noting that the doctrine of informed consent does not apply to ‘contractual arrangements in fields other than medicine; where caveat emptor remains the guide’” (Gray 1978: 40).

10. Some will argue that this preliminary testing on animals is itself unethical, but for present purposes we can table that very interesting issue.

11. Note that microlending programs have progressed *pari passu* with the new experimental development economics, and all the key ingredients are at play here: an initiative that has public benefit, science that satisfies the “facts-based” project of development, and, another undiscussed benefit, opportunity for profit, such as those made by the Grameen Bank, whose founder, Muhammad Yunus, won the Nobel Peace Prize in 2006.

12. An experimental intervention that had widespread benefits to participants and the community at large can be found in a now-famous paper by Miguel and Kremer (2004). The authors found that giving deworming drugs to children in some schools reduced the overall incidence of intestinal worms in the surrounding areas. In the parlance of economics, the deworming intervention was accompanied by a positive externality in that many who did not bear the cost of the treatment nevertheless gained from it. The benefit of the intervention, therefore, accrued both to subjects and to the community. But this may not always be the case.

13. Paluck and Green’s appendix I also discusses their procedure for informed consent (verbal) and the details surrounding their use of covert observation, where the behavior of participants was recorded after the experiment was supposedly over. The authors write, “Widely adopted ethical standards for IRBs state that recording behaviors anonymously (without recording the names of people enacting the behavior) does not require informed consent or debriefing” (2009: 39). I should point out that IRBs are *legal* boards that are primarily designed to shield the researcher and the university from legal entanglements. Many do have designs to ethical considerations, but my discussion from above, which highlights the respect for human dignity and autonomy, would preclude deliberately deceiving people about when their behavior was being recorded, regardless of the anonymity of the data.

14. To this end the authors procured funding to serve as an “active control” of counseling in case such trauma surfaces.

15. One would typically use the term *general equilibrium effect* to describe the influence that a shock in one part of an economic system has on other elements of that system, but, as Banerjee and Duflo (chapter 4 in this book) point out, economists think of multiple markets when they conceive of a general equilibrium. Since the effects we are interested with field experiments might not be in a separate market per se but in the changing social dynamics of a community, the authors prefer the term *equilibrium effects*. It would be even more precise to think of experiments as causing disequilibrium effects, in that experimental interventions are designed to take

relationships and structures that are, as it were, in some sort of equilibrium and upset them in order to see how people respond.

16. The principle of justice is therefore deliberately meant to discourage situations like the one in which rural men of color in Tuskegee, Alabama, are used as subjects in a study whose results might never be available to members of their communities (let alone to themselves).

17. In written comments to the author, Elizabeth Wood noted that the justice principle is sometimes interpreted to mean “that certain populations should not be excluded from experiments, and thus from the potential benefits (a critique that arose from gendered and racial bias in some medical experiments).” This is an interesting argument; it might be said that the “benefits” of research accrue to women and children, often the target populations of this research.

18. Another striking feature of the above quotation is that villagers are told that randomization is a “fair way” to allocate resources, when in fact resources are only allocated randomly so as to gain scientific leverage over the implemented program. At the most basic level, misrepresentation of the purposes of randomization as serving fundamental fairness rather than scientific measurement raises concerns within the principle of justice, which states that “equals ought to be treated equally” (*Belmont Report*: 6). In other words, participants should not be deliberately misled for reasons that are outside the research question.

19. The use of the word *tropics* here is an allusion to Deaton (2009), meant to highlight the bad taste that remains after confronting these types of arguments.

20. This is not to say that a similar problem does not arise when research is conducted on vulnerable populations in the United States. It does. I focus on developing countries because the locus of experimental work has shifted toward these areas, though there has been much discussion in the sociological literature already about this type of work in the United States.