Handbook of Field Experiments
The last 15 years have seen an explosion in the number, scope, quality, and creativity of field experiments. To take stock of this remarkable progress, we were invited to edit a Handbook of Field Experiments, published at Elsevier. We were fortunate to assemble a volume made of wonderful papers by the best experts in the field. Some chapters are more methodological, while others are focused on results. All of them provide thoughtful reflections on the advances and issues in the field, useful research tips and insights into what the next steps need to be, all of which should be very useful for graduate students. Taken together, these papers offer an incredibly rich overview of the state of literature. This page collects together all the working paper versions of the chapters, and will also link to the final versions as they become available. We hope you enjoy it.

—Abhijit Banerjee and Esther Duflo

Introduction

An Introduction to the "Handbook of Field Experiments"
Abhijit Banerjee and Esther Duflo

Many (though by no means all) of the questions that economists and policymakers ask themselves are causal in nature: What would be the impact of adding computers in classrooms? What is the price elasticity of demand for preventive health products? Would increasing interest
rates lead to an increase in default rates? Decades ago, the statistician Fisher (Fisher, 1925) proposed a method to answer such causal questions: Randomized Controlled Trials (RCTs). In an RCT, the assignment of different units to different treatment groups is chosen randomly. This ensures that no unobservable characteristics of the units are reflected in the assignment, and hence that any difference between treatment and control units reflects the impact of the treatment. While the idea is simple, the implementation in the field can be more involved, and it took some time before randomization was considered to be a practical tool for answering questions in economics.

Some Historical Background

The Politics and Practice of Social Experiments: Seeds of a Revolution
Judy Gueron

Between 1970 and the early 2000s, there was a revolution in support for the use of randomized experiments to evaluate social programs. Focusing on the welfare reform studies that helped to speed that transformation in the United States, this chapter describes the major challenges to randomized controlled trials (RCTs), how they emerged and were overcome, and how initial conclusions about conditions necessary to success — strong financial incentives, tight operational control, and small scale — proved to be wrong. The final section discusses lessons from this experience for other fields.
Randomized experiments have a long tradition in agricultural and biomedical settings. In economics they have a much shorter history. Although there have been notable experiments over the years, such as the RAND health care experiment (Manning, Newhouse, Duan, Keeler and Leibowitz, 1987, see the general discussion in Rothstein and von Wachter, 2016) and the Negative Income Tax experiments (e.g., Robins, 1985), it is only recently that there has been a large number of randomized experiments in economics, and development economics in particular. See Duflo, Glennerster, and Kremer (2006) for a survey. In this chapter we discuss some of the statistical methods that are important for the analysis and design of randomized experiments. A major theme of the chapter is the focus on statistical methods directly justified by randomization, in the spirit of Freedman who wrote “Experiments should be analyzed as experiments, not as observational studies. A simple comparison of rates might be just the right tool, with little value added by ‘sophisticated’ models,” (Freedman, 2006, p. 691) We draw from a variety of literatures. This includes the statistical literature on the analysis and design of experiments, e.g., Wu and Hamada (2009), Cox and Reid (2000), Altman (1991), Cook and DeMets (2008), Kempthorne (1952, 1955), Cochran and Cox (1957), Davies (1954), and Hinkelman and Kempthorne (2005, 2008). We also draw on the literature on causal inference, both in experimental and observational settings, Rosenbaum (1995, 2002, 2009), Rubin (2006), Cox (1992), Morgan and Winship (2007), Morton
Williams (2010) and Lee (2005), and Imbens and Rubin (2015). In the economics literature we build on recent guides to practice in randomized experiments in development economics, e.g., Duflo, Glennerster, and Kremer (2006), Glennerster (2016), and Glennerster and Takavarasha (2013) as well as the general empirical micro literature (Angrist and Pischke, 2008).

**Decision Theoretic Approaches to Experiment Design and External Validity**

Abhijit Banerjee, Sylvain Chassang, and Erik Snowberg

A modern, decision-theoretic framework can help clarify important practical questions of experimental design. Building on our recent work, this chapter begins by summarizing our framework for understanding the goals of experimenters, and applying this to re-randomization. We then use this framework to shed light on questions related to experimental registries, pre-analysis plans, and most importantly, external validity. Our framework implies that even when large samples can be collected, external decisionmaking remains inherently subjective. We embrace this conclusion, and argue that in order to improve external validity, experimental research needs to create a space for structured speculation.

**The Practicalities of Running Randomized Evaluations: Partnerships, Measurement, Ethics, and Transparency**

Rachel Glennerster

Economists have known for a long time that randomization could help identify causal connections by solving the problem of selection bias. Chapter 1 in this book and Gueron and Rolston (2013) describe the effort in the US to move experiments out of the laboratory into the policy world in the
1960s and 1970s. This experience was critical in proving the feasibility of field experiments, working through some of the important ethical questions involved, showing how researchers and practitioners could work together, and demonstrating that the results of field experiments were often very different from those generated by observational studies. Interestingly, there was relatively limited academic support for this first wave of field experiments (Gueron and Rolston 2013), most of which were carried out by research groups such as MDRC, Abt, and Mathematica, to evaluate US government programs, and they primarily used individual-level randomization. In contrast, a more recent wave of field experiments starting in the mid-1990s was driven by academics, initially was focused on developing countries, often worked with nongovernmental organizations, and frequently used clustered designs.

The Psychology of Construal in the Design of Field Experiments
Elizabeth Levy Paluck and Eldar Shafir

Why might you be interested in this chapter? A fair assumption is that you are reading because you care about good experimental design. To create strong experimental designs that test people's responses to an intervention, researchers typically consider the classically recognized motivations presumed to drive human behavior. It does not take extensive psychological training to recognize that several types of motivations could affect an individual's engagement with and honesty during your experimental paradigm. Such motivations include strategic self-presentation, suspicion, lack of trust, level of education or mastery, and simple utilitarian motives such as least effort and optimization. For example, minimizing the extent to which your findings are
attributable to high levels of suspicion among participants, or to their decision to do the least amount possible, is important for increasing the generalizability and reliability of your results.

Understanding Preferences and Preference Change

Field Experiments in Markets
Omar Al-Ubaydli and John List
This is a review of the literature of field experimental studies of markets. The main results covered by the review are as follows: (1) Generally speaking, markets organize the efficient exchange of commodities; (2) There are some behavioral anomalies that impede efficient exchange; (3) Many behavioral anomalies disappear when traders are experienced.

Field Experiments on Discrimination
Marianne Bertrand and Esther Duflo
This article reviews the existing field experimentation literature on the prevalence of discrimination, the consequences of such discrimination, and possible approaches to undermine it. We highlight key gaps in the literature and ripe opportunities for future field work. Section 1 reviews the various experimental methods that have been employed to measure the prevalence of discrimination, most notably audit and correspondence studies; it also describes several other measurement tools commonly used in lab-based work that deserve greater consideration in field research. Section 2 provides an overview of the literature on the costs of being stereotyped or discriminated against, with
a focus on self-expectancy effects and self-fulfilling prophecies; section 2 also discusses the thin field-based literature on the consequences of limited diversity in organizations and groups. The final section of the paper, Section 3, reviews the evidence for policies and interventions aimed at weakening discrimination, covering role model and intergroup contact effects, as well as socio-cognitive and technological de-biasing strategies.

Field Experiments on Voter Mobilization: An Overview of a Bourgeoning Literature
Alan Gerber and Donald Green

In recent years the focus of empirical work in political science has begun to shift from description to an increasing emphasis on the credible estimation of causal effects. A key feature of this change has been the increasing prominence of experimental methods, and especially field experiments. In this chapter we review the use of field experiments to study political participation. Although several important experiments address political phenomena other than voter participation (Bergan 2009; Butler and Broockman 2015; Butler and Nickerson 2011; Broockman 2013, 2014; Grose 2014), the literature measuring the effect of various interventions on voter turnout is the largest and most fully developed, and it provides a good illustration of how the use of field experiments in political science has proceeded. From an initial focus on the relative effects of different modes of communication, scholars began to explore how theoretical insights from social psychology and behavioral economics might be used to craft messages and how voter mobilization experiments could be employed to test the real world effects of theoretical claims. The existence of a large number of experimental turnout studies was essential, because it
provided the background against which unusual and important results could be easily discerned.

**Lab in the Field: Measuring Preferences in the Wild**

Uri Gneezy and Alex Imas

In this chapter, we discuss the “lab-in-the-field” methodology, which combines elements of both lab and field experiments in using standardized, validated paradigms from the lab in targeting relevant populations in naturalistic settings. We begin by examining how the methodology has been used to test economic models with populations of theoretical interest. Next, we outline how lab-in-the-field studies can be used to complement traditional Randomized Control Trials in collecting covariates to test theoretical predictions and explore behavioral mechanisms. We proceed to discuss how the methodology can be utilized to compare behavior across cultures and contexts, and test for the external validity of results obtained in the lab. The chapter concludes with an overview of lessons on how to use the methodology effectively.

**Field Experiments in Marketing**

Duncan Simester

Marketing is a diverse field that draws from a rich array of disciplines and a broad assortment of empirical and theoretical methods. One of those disciplines is economics and one of the methods used to investigate economic questions is field experiments. The history of field experiments in the marketing literature is surprisingly long. Early examples include Curhan (1974) and Eskin and Baron (1977), who vary prices, newspaper advertising, and display variables in grocery stores. This chapter reviews the recent history of field experiments in marketing by identifying
papers published in the last 20 years (between 1995 and 2014). We report how the number of papers published has increased during this period, and evaluate different explanations for this increase. We then group the papers into five topics and review the papers by topic. The chapter concludes by reflecting on the design of field experiments used in marketing, and proposing topics for future research.

The Challenge of Improving Human Capital

Impacts and Determinants of Health Levels in Low-Income Countries
Pascaline Dupas and Ted Miguel

Improved health in low-income countries could considerably improve wellbeing and possibly promote economic growth. The last decade has seen a surge in field experiments designed to understand the barriers that households and governments face in investing in health and how these barriers can be overcome, and to assess the impacts of subsequent health gains. This chapter first discusses the methodological pitfalls that field experiments in the health sector are particularly susceptible to, then reviews the evidence that rigorous field experiments have generated so far. While the link from in utero and child health to later outcomes has increasingly been established, few experiments have estimated the impacts of health on contemporaneous productivity among adults, and few experiments have explored the potential for infrastructural programs to impact health outcomes. Many more studies have examined the determinants of individual health behavior, on the side of consumers as well as among providers of health products and services.
The Production of Human Capital in Developed Countries: Evidence from 196 Randomized Field Experiments

Roland Fryer

Randomized field experiments designed to better understand the production of human capital have increased exponentially over the past several decades. This chapter summarizes what we have learned about various partial derivatives of the human capital production function, what important partial derivatives are left to be estimated, and what – together – our collective efforts have taught us about how to produce human capital in developed countries. The chapter concludes with a back of the envelope simulation of how much of the racial wage gap in America might be accounted for if human capital policy focused on best practices gleaned from randomized field experiments.

Field Experiments in Education in Developing Countries

Karthik Muralidharan

Perhaps no field in development economics in the past decade has benefited as much from the use of experimental methods as the economics of education. The rapid growth in high-quality studies on education in developing countries (many of which use randomized experiments) is perhaps best highlighted by noting that there have been several systematic reviews of this evidence aiming to synthesize findings for research and policy in just the past three years. These include Muralidharan 2013 (focused on India), Glewwe et al. 2014 (focused on school inputs), Kremer et al. 2013, Krishnaratne et al. 2013, Conn 2014 (focused on sub-Saharan Africa), McEwan 2014, Ganimian and Murnane (2016), Evans and Popova (2015), and Glewwe and Muralidharan (2016). While these are not all restricted to experimental studies, they
typically provide greater weight to evidence from randomized controlled trials (RCT's).

Designing Effective Social Programs

Social Policy: Mechanism Experiments and Policy Evaluations
Bill Congdon, Jeffrey Kling, Jens Ludwig, and Sendhil Mullainathan

Policymakers and researchers are increasingly interested in using experimental methods to inform the design of social policy. The most common approach, at least in developed countries, is to carry out large-scale randomized trials of the policies of interest, or what we call here policy evaluations. In this chapter we argue that in some circumstances the best way to generate information about the policy of interest may be to test an intervention that is different from the policy being considered, but which can shed light on one or more key mechanisms through which that policy may operate. What we call mechanism experiments can help address the key external validity challenge that confronts all policy-oriented work in two ways. First, mechanism experiments sometimes generate more policy-relevant information per dollar of research funding than can policy evaluations, which in turn makes it more feasible to test how interventions work in different contexts. Second, mechanism experiments can also help improve our ability to forecast effects by learning more about the way in which local context moderates policy effects, or expand the set of policies for which we can forecast effects. We discuss how mechanism experiments and policy evaluations can complement one another, and provide examples from a range of social policy areas including health insurance, education, labor market
policy, savings and retirement, housing, criminal justice, redistribution, and tax policy. Examples focus on the U.S. context.

Field Experiments in Developing Country Agriculture
Alain de Janvry, Elisabeth Sadoulet, and Tavneet Suri
This chapter provides a review of the role of field experiments in answering research questions in agriculture that ultimately let us better understand how policy can improve productivity and farmer welfare in developing economies. We first review recent field experiments in this area, highlighting the contributions experiments have already made to this area of research. We then outline areas where experiments can further fill existing gaps in our knowledge on agriculture and how future experiments can address the specific complexities in agriculture.

The Personnel Economics of the State
Frederico Finan, Ben Olken, and Rohini Pande
Governments play a central role in facilitating economic development. Yet while economists have long emphasized the importance of government quality, historically they have paid less attention to the internal workings of the state and the individuals who provide the public services. This chapter reviews a nascent but growing body of field experiments that explores the personnel economics of the state. To place the experimental findings in context, we begin by documenting some stylized facts about how public sector employment differs from that in the private sector. In particular, we show that in most countries throughout the world, public sector employees enjoy a significant wage premium over their private sector counterparts. Moreover, this wage gap is largest among low-income countries, which tends to be
precisely where governance issues are most severe. These differences in pay, together with significant information asymmetries within government organizations in low-income countries, provide a prima facie rationale for the emphasis of the recent field experiments on three aspects of the state–employee relationship: selection, incentive structures, and monitoring. We review the findings on all three dimensions and then conclude this survey with directions for future research.

**Designing Social Protection Programs: Using Theory and Experimentation to Understand how to Help Combat Poverty**

Rema Hanna and Dean Karlan

“Anti-poverty” programs come in many varieties, ranging from multi-faceted, complex programs to more simple cash transfers. Articulating and understanding the root problem motivating government and nongovernmental organization intervention is critical for choosing amongst many anti-poverty policies, or combinations thereof. Policies should differ depending on whether the underlying problem is about uninsured shocks, liquidity constraints, information failures, or some combination of all of the above. Experimental designs and thoughtful data collection can help diagnose the root problems better, thus providing better predictions for what anti-poverty programs to employ in specific conditions and contexts. However, the more complex theories are likewise more challenging to test, requiring larger samples, and often more nuanced experimental designs, as well as detailed data on many aspects of household and community behavior and outcomes. We provide guidance on these design and testing issues for social protection programs, from how to target programs, to who should implement the program, to whether and what conditions to require for
program participation. In short, careful experimentation designed testing can help provide a stronger conceptual understanding of why programs do or not work, thereby allowing one to ultimately make stronger policy prescriptions that further the goal of poverty reduction.

Social Experiments in the Labor Market
Jesse Rothstein and Till von Wachter

Large-scale social experiments were pioneered in labor economics, and are the basis for much of what we know about topics ranging from the effect of job training to incentives for job search to labor supply responses to taxation. Random assignment has provided a powerful solution to selection problems that bedevil non-experimental research. Nevertheless, many important questions about these topics require going beyond random assignment. This applies to questions pertaining to both internal and external validity, and includes effects on endogenously observed outcomes, such as wages and hours; spillover effects; site effects; heterogeneity in treatment effects; multiple and hidden treatments; and the mechanisms producing treatment effects. In this Chapter, we review the value and limitations of randomized social experiments in the labor market, with an emphasis on these design issues and approaches to addressing them. These approaches expand the range of questions that can be answered using experiments by combining experimental variation with econometric or theoretical assumptions. We also discuss efforts to build the means of answering these types of questions into the ex ante design of experiments. Our discussion yields an overview of the expanding toolkit available to experimental researchers.