ROUTLEDGE STUDIES IN EXPERIMENTAL POLITICAL SCIENCE

Field Experiments in Political Science and Public Policy

Practical Lessons in Design and Delivery

Peter John

ROUTLEDGE

In this lively and accessible book, Peter John has shared with readers a wealth of practical guidance that comes from years and years of field experimentation in political science and public policy. Every researcher should heed its advice before venturing into the field.

Donald P. Green, Columbia University, USA

Experiments are revolutionizing what we think we know about many aspects of politics. This volume contributes nicely to this body of work by extending the focus to critical questions of public policy. This is a timely and important book.

Costas Panagopoulos, Fordham University, USA

This page intentionally left blank

Field Experiments in Political Science and Public Policy

Field experiments – randomized controlled trials – have become ever more popular in political science, as well as in other disciplines, such as economics, social policy, and development. Policy-makers have also increasingly used randomization to evaluate public policies, designing trials of tax reminders, welfare policies, and international aid programs to name just a few of the interventions tested in this way. The reason experiments have become so successful is because they assess causal claims in ways that other methods of evaluation find hard to emulate.

Social scientists and evaluators have rediscovered how to design and analyze field experiments, but they have paid much less attention to the challenges of organizing and managing them. Trials pose unique challenges and opportunities for the researcher and evaluator which come from working in the field. The research experience can be challenging and at times hard to predict. This book aims to help researchers and evaluators plan and manage their field experiments so they can avoid common pitfalls. It is also intended to open up discussion about the context and backdrop to field experiments so that these practical aspects of field experiments are better understood.

The book sets out ten steps researchers can use to plan their field experiments, then nine threats to watch out for when they implement them. There are case studies of voting and political participation, elites, welfare and employment, nudging citizens, and developing countries.

Peter John is Professor of Political Science and Public Policy in the Department of Political Science, University College London. He is an expert on the theory and practice of public policy, and has pioneered the use of field experiments in the study of civic participation in the UK. He is an academic advisor to the Behavioural Insights Team.

Routledge Studies in Experimental Political Science
Edited by Kevin Arceneaux, Temple University and
Cindy D. Kam, Vanderbilt University

Experimental methods are now firmly established within political science and are widely employed across its diverse empirical subfields to study important political phenomena. The logic of experimentation makes it an appealing and powerful methodological tool that enables scholars to establish causality and probe into the mechanisms underlying observable regularities. Experiments, because of their transparency, also enable researchers to communicate their findings to a broad audience. Although highly technical knowledge is not necessary for understanding the gist of experiments, experiments must be designed, administered, and analyzed with care and attention to detail.

The *Routledge Studies in Experimental Political Science* series was developed to publish books that educate readers about the appropriate design and interpretation of experiments and books that showcase innovative and important applications of experimental work.

1. Ethics and Experiments
Problems and Solutions for Social Scientists and Policy Professionals
Edited by Scott Desposato

2. Field Experiments in Political Science and Public Policy
Practical Lessons in Design and Delivery
Peter John

Field Experiments in Political Science and Public Policy

Practical Lessons in Design and Delivery

Peter John

Routledge
Taylor & Francis Group
NEW YORK AND LONDON

First published 2017
by Routledge
711 Third Avenue, New York, NY 10017

and by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

Routledge is an imprint of the Taylor & Francis Group, an informa business

Library of Congress Cataloging in Publication Data
Names: John, Peter, 1960- author.
Title: Experimentation in political science and public policy : practical lessons for the delivery of randomized experiments / Peter John.
Description: New York, NY : Routledge is an imprint of the Taylor & Francis Group, an Informa Business, [2017] | Series: Routledge studies in experimental political science ; 2 | Includes bibliographical references and index.
Identifiers: LCCN 2016032526| ISBN 9781138776821 (hbk) | ISBN 9781138776838 (pbk)
Subjects: LCSH: Political science–Research–Methodology.
Classification: LCC JA86 .J58 2017 | DDC 320.072/4–dc23
LC record available at https://lccn.loc.gov/2016032526

ISBN: 978-1-138-77682-1 (hbk)
ISBN: 978-1-138-77683-8 (pbk)
ISBN: 978-1-315-77302-5 (ebk)

Typeset in Bembo
by Taylor & Francis Books

Contents

	Series Editors' Foreword	viii
	Preface	ix
	Acknowledgments	xi
1	Field Experimentation: Opportunities and Constraints	1
2	Essential Steps for the Design of Field Experiments	17
3	Implementing Field Experiments	49
4	A Brief History of Field Experimentation	63
5	Experiments on Voting and Political Participation	74
6	Experiments on Elites	95
7	Experiments in Welfare and Employment	110
8	Nudging Citizens	134
9	Field Experiments in Developing Countries	153
10	Conclusion	173
	References	180
	Index	198

Series Editors' Foreword

One of the first political science experiments was a field experiment conducted by Harold Gosnell in 1924. Gosnell stepped out of his University of Chicago office and conducted an experiment in the real world – in the neighborhoods of Chicago – to investigate whether encouraging people to vote actually pushed them to the polls. Over the past few decades, political scientists have conducted thousands of field experiments to study a diverse range of topics, including Gosnell's subject of voter turnout as well as such subjects as public policy effectiveness and the responsiveness of elected officials. Political scientists have implemented field experiments not only in their own backyards but also, and increasingly, worldwide. While a number of existing methodological resources provide guidance on initial design and data analysis, many critical steps transpire in between the initial design and subsequent analysis. There are comparatively fewer how-to resources that offer useful advice for this decisive stage of administering field experiments.

Peter John's insightful and highly useful monograph provides the researcher with important information in successfully implementing a field experiment. Field experiments typically require coordination among multiple individuals, organizations, and institutions to design and administer treatments and record outcomes. As John notes, field experimentalists often confront many potential pitfalls that could invalidate the time, effort, and financial resources dedicated to the experiment. Drawing upon decades of his own experience and interviews with scholars across a variety of research domains, John provides scholars with insights into what works, what can go wrong, and what pitfalls can be prevented in field experiments. John distills these experiences into ten essential steps to undertake before administering the experiment and nine problems to avoid once the study is in the field. He illustrates these principles of practice across an array of research areas in political science and public policy. In doing so, this monograph offers practical and concrete advice to just about anyone in political science planning to undertake a field experiment.

Kevin Arceneaux, Temple University
Cindy D. Kam, Vanderbilt University

Preface

I carried out my first field experiment in 1999. It was a fairly modest intervention that tested whether a visit by young people to the Houses of Parliament and No. 10 Downing Street would change their knowledge and attitudes towards politics compared to those who did not get this opportunity (Halpern and John 2005). In spite of only shifting political knowledge in a very small positive direction, I was intrigued by this project, not least by the challenge of getting a debate going between the young people and the prime minister's education advisor at time, Andrew Adonis. When I arrived at the University of Manchester in 2004, experiments came up again as a method that the Institute for Political and Economic Governance wanted to deploy in its evaluations, so I decided to try my hand at one. This experiment was a Get Out the Vote intervention carried out in the constituency of Wythenshawe and Sale East just before the 2005 General Election (John and Brannan 2008). From then on, I was hooked and I have been doing experiments in earnest ever since. They have come to dominate my life, and I usually start about three or four each year, sometimes more.

In this book, I wish to convey my engagement and support for the method, which might appear to be a very technical subject, one for the statistician, so rather dry at first brush. To be sure, there are many technical issues to resolve that are intriguing and fascinating in their own right as well as essential when doing an experiment. But I cannot say enough about the pleasure of designing and carrying out experiments as a research experience. Of course, I have felt intense frustration when implementing some experiments, and every experimenter has to cope with both the highs and lows on getting the results back. Yet, as soon as the last experiment is over, I am up for the next one no matter how tough it has been.

When I read the outstanding introductions and exemplary peer-reviewed articles that indicate the state of the art in social science today, I do not usually find out about the features of experiments that excite and challenge me. Many published studies do not fully inform the reader about what exactly happened during an experiment. In the published paper, the complicated story gets compressed into a terse design section. I suspect that the reason there are so many footnotes containing these practical details is

because puzzled reviewers often ask for them. When attending presentations of experiments in academic gatherings, I often observe participants asking many questions about design and delivery, wanting to know about the choices the researcher made. The purpose of these interrogations is to find out whether the experiment was delivered effectively and to reassure other researchers that it does not display bias in its findings. With these considerations in mind, this book aims to illuminate the design challenges and experiences of implementing experiments. I draw on my experiences and those of others to explain how to carry out effective field experiments. The hope is that researchers will pay more attention to these issues when planning experiments and also write more about the decisions they made about design and implementation.

Field experimenters in political science and public policy need to pay particular attention to research design because of the special features of their method: a trial usually needs the involvement of practitioners and public agencies to get the intervention implemented; and it requires precise control over a range of practical tasks that enfold over time. This task is a much greater burden for a researcher or policy-maker than is required in an observational study, because of the need to supervise or facilitate an intervention at the same time as collecting and analyzing data. These two activities need to be linked together and coordinated in a way that respects the underlying assumptions of the trial. To get an experiment off the ground, a researcher uses much unstated knowledge as well as a lot of common sense. Most experimenters learn their skills in the field, often the hard way by making mistakes and correcting them during the research process. If I can help students and other researchers avoid some unnecessary errors, I will consider my job well done, although unexpected snags always crop up just when least expected and no one can prepare for every eventuality. In this way, the book project extends my role as a teacher and an advocate. It is very pleasing to be able to summarize the lessons and skills and I have learned and recount others that I have discovered through discussions with experts in the field.

Acknowledgments

I have many people and organizations to thank. First is my publisher, Routledge/Taylor & Francis. Craig Fowlie very much wanted me to do this book and successfully persuaded me to sign up. As ever, I appreciate his enthusiasm and patience. No less important is Natalja Mortensen, my US editor, who has managed the project and kept me motivated. I thank her and other staff at Routledge, in particular Lillian Rand and Paola Celli. I also would like to show my gratitude to the series editors Kevin Arceneaux and Cindy Kam for handling the proposal so professionally and for their extensive comments on each draft, which improved the manuscript immensely. The reviewers were very sympathetic to the book and made very useful suggestions.

I also thank the Wagner School of Public Policy, which housed me in October and November of 2014, giving me the space to start to write the book and the facility to approach the many experimentalists who work in the New York region. My friend and collaborator Tony Bertelli was associate dean there, and he assisted my stay by finding me an office and looked after me with his characteristic solicitude. I am also appreciative to the Department of Political Science, University College London for giving a sabbatical term, which permitted the visit to New York as well as writing time at home.

I have learned a lot from other people about experiments over the years, first from David Halpern when we were researching together on a project on social capital. I then benefited from the generosity of Don Green who helped me with my first voting experiment in 2005 and has been a diligent advisor to my experiments ever since. I am also grateful to colleagues with whom I have designed and implemented experiments: Toby Blume, Tessa Brannan, Florian Foos, Sarah Cotterill, Hanhua Liu, Helen Margetts, Alice Moseley, Liz Richardson, Matt Ryan, Julie Van de Vyver, Michael Sanders, Antonio Silva, and Gerry Stoker. Experimenting works best when done collaboratively, and I have learned a great deal from these individuals when delivering trials. I have found that the officials and employees of organizations I have worked with have taught me a lot about how public administration works today and helped me to understand how to best deliver a field

experiment in a complex environment. I thank these people as they have gone out of their way to do high-quality evaluations while doing their hard-pressed jobs at the same time. I have also been privileged to be a member of the network, Evidence in Government and Politics (EGAP). Its members tolerated me as someone from outside the world of development experiments, and I learned a lot from going to its meetings and talking to participants, some of whom were interviewed for Chapter 9. Similarly, I have been honored to be a member of the Academic Advisory Panel of the Behavioural Insights Team, which has allowed me to observe and sometimes participate in some of the many experiments the team has carried out since 2010.

My first attempt to set out my approach to experimentation was at my inaugural lecture at University College London on November 20, 2012 (John 2013), and I appreciate members of my department and guests for listening so patiently (these lectures forbid the posing of questions). I also had a chance to try out some of the ideas in the book during the short course, "Field Experiments and Randomized Controlled Trials," which I taught at the University of Essex from April 29 to May 1, 2014. I organized the course around a series of steps needed to design a trial. As the format worked, I used it to structure the materials in Chapter 2. The participants also alerted me to the need for practical advice on how to design and carry out field experiments, which they said was not available in the books on the subject, motivating me to persist with the theme and approach of this book. I thank the students for their patience with my teaching and for communicating enthusiasm for their own field experiments. With Toby Blume, I have introduced practitioners to trials in a series of workshops also using the steps approach. We got participants to fill out a form containing boxes for each step. I thank both Toby and local authority officers for helping me to think about the practical side of the design of trials while running these sessions. I also presented the ideas behind this book during a short trip to Switzerland in early December 2014 when I visited the universities of Lausanne and St Gallen. I am very grateful to participants at these seminars for their engagement with the project, and again it was the research students who responded well to the materials, telling me that I was on the right track.

As part of the preparation for the book, I have approached a number of leading scholars and researchers in the field who have either been interviewed or have agreed to talk more informally on this subject. Interviewees were Sheree Bennett, Chris Blattman, Dan Butler, Saguara Datta, Don Green, Christian Grose, Macartan Humphreys, Gwyneth McClendon, Valérie-Anne Mahéo, David Nickerson, Ana de la O, and Costas Panagopoulos. More informal discussions have been had with Karen Grépin, Todd Rogers, Cyrus Sami, Jonathan Morduch, and Becky Morton. I had two great meetings at MDRC in New York, and I am very grateful to Jim Riccio for his generosity in setting them up and in taking the time to speak to me.

Once I had produced a very rough draft of the book, Don Green, Florian Foos, and Manu Savani read it thoroughly and patiently pointed out its many errors, as well as making many helpful suggestions for which I am very grateful. Finally, Rebecca Pizzitola helped me to produce a fair copy of the draft and I thank her too. It is always said in acknowledgments of this kind that anyone who had anything to do with the book has no responsibility for its contents. All I can say is that this statement is true with this book, and I can't stress its importance enough.

This page intentionally left blank

1 Field Experimentation: Opportunities and Constraints

Introduction

Field experiments have come back into fashion in political science, as well as in other disciplines, such as economics, social policy, and development, with social scientists using randomized controlled trials very much more than they did previously. Policy-makers have also increasingly used randomization to get better traction on the evaluation of policies, designing trials of tax reminders, welfare policies, and international aid programs to name just a few of the interventions tested in this way. Social scientists and evaluators have rediscovered how to design and analyze field experiments, but they have paid much less attention to the challenges of organizing and managing them, and in understanding the overall context of the research process. Field experiments pose unique challenges and opportunities for the researcher and evaluator, who want to know how to do them efficiently and get the most from them. These challenges often come from the constraints of working in the real world, the field element of experiments, which makes the research experience challenging and at times hard to predict. This book is aimed to help researchers and evaluators plan and manage their field experiments in the light of the practical constraints and contingencies that emerge. It is also intended to open up discussion about the context and backdrop to field experiments so that these practical aspects of field experiments are better understood and also written up in more detail in published studies.

This chapter starts off defining the key terms, explaining what an experiment is. It sets out why a randomized experiment is a preferred method for social scientists and policy-makers. It then goes on to explain why the practical side to a field experiment is so important. There is an example to show how an experiment operates in the field. The rest of the chapter sets out the main argument and plan of the book.

What is an experiment?

An experiment occurs when human beings manipulate the world to understand causal relationships. It is a common means by which scientists

acquire knowledge. The secret is to have precise control over the different elements of an intervention and to measure carefully what happens during the experiment. It is this degree of control over the external world that is the big attraction of experiments, which explains why researchers from other disciplines seek to emulate the method. By intervening as well as observing an experiment unlocks knowledge in ways that cannot be fully achieved by observation alone.

In the social sciences and in government, the researcher or policy-maker looks for or creates some random or chance variation that ensures that the difference in outcomes is only associated with an intervention or policy compared to the status quo. In what is called a natural experiment, accidental differences between populations or areas that have been created by government or by nature, such as the impact of boundaries or cut-offs for eligibility in a policy program, may be thought of as if they were random (see Dunning 2012). Sometimes randomness happens directly, such as in a government decision or institutional rule that can be evaluated like an experiment. For example, when the US government drafted young people to the Vietnam War, they selected certain starting numbers related to birth date. Erikson and Stoker (2011) used this random allocation to find out whether people who served in the war were more liberal as a result and had different political behaviors from those that did not (they did).

More often, researchers and policy-makers create the random variation themselves. This method is sometimes called artificial randomization, but is more commonly known as the randomized controlled trial, where individuals or communities or other units are randomly allocated to two or more groups. In this procedure, one or more groups get an intervention while another group, the control, does not. A comparison can be made between the outcomes in these groups to see if the intervention (also commonly called the treatment) made an impact. So long as there are at least two comparison groups and one group receives an intervention while another does not – or gets a different intervention – and assignment to these groups is random, it is possible to make an inference that the difference in outcomes between the groups – other than random variation – has only been caused by the intervention.

This kind of method is also called a field experiment, which is done in the community at large or within an organization. It may be contrasted with a laboratory experiment done in controlled settings (excepting the hybrid form called "lab in the field" where a laboratory experiment is done in the field). More generally, people use the term "trial" or the ubiquitous acronym, RCT. For example, if a public agency wants to test whether giving training gives those who are unemployed more of a chance to find employment, it can find a group of welfare recipients to try out an intervention; the agency can then randomly allocate the training to one group and leave the other alone or just provide a normal level of service. The agency can then find out if those who got the training were more likely to find employment than those who did not.

The design of a field experiment

There are four main stages to a field experiment. First is the selection of the sample, which involves finding the relevant population from which the researcher can answer the research question or questions. Second, the researcher or policy-maker randomizes the units of the sample – for instance, people – into groups. Third, the policy, program, or intervention is applied to one group, leaving the other group alone or just with a normal service or state of affairs, or with another program. Fourth, and finally, outcomes are measured after the intervention has taken place (which can also be done at the start to compare change over time). In this outline, the design is quite simple, which is one of the attractions of a field experiment. The method is quite easy to explain to a layperson, practitioner, or student in spite of its cumbersome name. A practitioner or student grasps the idea that one group is randomly chosen to be "policy on" and another is "policy off"; and that outcomes are compared between two or more groups to find out whether the policy worked or not. Because of the success of the trial in medical evaluations and because examples are extensively reported in the media, most people have a good conception of what one is: a policy or social science trial is equivalent to testing a new medicine on a randomly allocated group of volunteers whose health outcomes are compared with those who do not get the medicine or get a placebo.

Of course, a field experiment can get complicated very quickly, not least in its design. It is possible to have many intervention groups that allow for multiple or several comparisons. Normally, randomization takes place across people or other units at the start of a trial, but a trial can have several stages with further randomizations that create new intervention groups or randomize the same people into different groups over time. Further, there are different kinds of randomization, such as when people or units are sorted into pairs and one person or unit out of each pair is picked randomly. Nonetheless, even with these variations, the idea of comparing outcomes across randomly allocated groups remains recognizable even in the most complex designs.

Making a causal inference

One of the most attractive features of a field experiment is that it can offer a clear test of a hypothesis, usually resulting in either a yes or no to the research question. For policy-makers, a well-designed field experiment can confirm whether an intervention worked or not. Moreover, if enough trials are done, it is possible to come to a conclusion that can be generalized across places and time periods. It is the claim to answer questions about causation that is the main appeal of a trial and explains why it has become so popular as a method in recent years.

Traditional social science methods

It might sound surprising that social science, which started using advanced statistical methods in earnest from the 1940s, should still want better leverage on causal questions. But it is very hard to make a causal inference in social research, even when there is a very strong theory about the determinants of a particular outcome and good measurements of what might cause it. This weakness occurs because it is not possible to rule out with certainty that factors other than the hypothesized one caused the outcome of interest or to confirm the direction of causation in a correlation. In social science as it has been practiced over the last 50 years or so, it is common to observe the simultaneous occurrence of an influencing and influenced variable and to measure its association. Take the example of canvassing a household by speaking to them face-to-face to persuade them to vote in an upcoming election – what is called a Get Out the Vote (GOTV) campaign. The research question is whether canvassing causes members of the household turn out to vote. There are various methods that could be used to test such a proposition. If there were data from a survey available that measured whether the respondents could recall having voted and having being canvassed, it would be possible to correlate or associate the two measures. It may be the case that the political parties keep records of which people they have canvassed that can be tied to the publically available electoral registers. It is very likely there is a positive relationship. However, it is not possible to rule out other factors that may cause people to turn out to vote are also correlated with being canvassed. The correlation between the two is just that and does not represent a causal relationship from canvassing to turnout.

Researchers who analyze observational data have been aware of the problem of establishing causation for a long time. They have developed a number of strategies to overcome it. One is to consider the possible correlates between the outcome and the intervention, collect data on each one, and then control for the association between the hypothesized variable and the outcome of interest. This strategy is achieved by multiple regression, where several variables, including the one of interest, are treated as independent causes of the outcome. If the association between the independent variable of interest and the outcome still remains as statistically significant after this procedure has been followed, it can be more reasonably concluded that there is a causal relationship. This inference is thought to be valid because other causal pathways have been allowed for and the item of interest – in this case canvassing – is still a factor determining the outcome. This strategy becomes more convincing if the researcher has considered in advance all possible causes of an outcome from theory and existing empirical work, measured and collected data on these alternative explanations and introduced the causal propositions as independent variables in the statistical model.

Yet, it may be possible that the researcher misses an alternative explanation or cannot accurately measure what causes the outcome, making the causal

inference hard to support even when control variables have been used. It is often said that the relationship is confounded or that there are confounders in play – that is, factors that are associated with both the intervention and the outcome being evaluated. The researcher can only guess what they are and cannot effectively control for them. There may also be other unobserved processes at work. Another problem is that the outcomes for individuals or units will vary over time. Their outcomes might get worse or better in ways that follow a natural cycle, such as finding a job, becoming healthier or the opposite. In this case, all the program or intervention picks up is what would have happened anyway. It does not have the counterfactual, which is the state of affairs that would have occurred in absence of the program. In all of these circumstances, it is important to have a method that rules out determinants of an outcome other than the intervention. Randomization generates the counterfactual of what would have happened without the intervention, such as between canvassed and non-canvassed individuals or households, or between unemployed people some of whom received a job training program.

The assumptions of a randomized controlled trial

Even though a trial can yield a causal inference, it can only do so if it is implemented correctly; in particular, an experiment should be delivered in a way that respects its assumptions. The starting point for an experiment is that it is not possible in any one individual to observe both the intervention or manipulation and thus the counterfactual. It is better to speak of "potential outcomes" that represent the outcomes for an individual had they received the treatment or control respectively. Because the counterfactual is not directly observed, it is necessary to build in assumptions when working out what causes the difference in outcomes between people or units that are in the treatment and control group (Rubin 1974). It is possible to work through these assumptions to show how a randomized controlled trial can give a true estimate of the effect of the treatment (for summaries, see Sekhon 2010; Deaton 2009; Gerber and Green 2012, 21–44). This formulation is what is referred to as causal inference or Rubin's (1974) causal model (RCM) or the Neyman–Rubin causal model.

The key assumptions are excludability – that the randomly assigned units only get the treatment and the control units do not – and non-interference – that randomly assigned units do not affect each other (sometimes called the Stable Unit Treatment Value Assumption, or SUTVA). By remembering these assumptions, the experienced researcher knows the trial design needs to have certain features – in particular, that random allocation should be respected throughout the implementation process. It is possible to watch out for whether members of the treatment and control groups make contact and influence each other (especially if there are parts of the treatment and control groups that are proximate to each other which allows for measurements to take place). Every trialist should keep careful watch of the flow of subjects or

people or areas through the experiment to ensure that there is no dropout or attrition, and to monitor it, so they know if one of the assumptions is broken and such data can be used to correct the results if needed.

If the basic design features are there and the protocol is followed, then not much should go wrong. If a field experiment becomes complicated, such as when people or units are lost to one of the intervention groups or to the control, the problem of one-sided attrition, the formal framework of potential outcomes can be helpful in deciding how to correct for such a violation of the assumptions and in finding out what technique to use to produce a valid estimate of the treatment.

An example for illustration

A lot of the above might sound rather abstract, so it is useful to talk about a practical example from political science, although the example could easily be from public policy. Even though there are many contexts across political science and public policy, and different problems to solve, the basic features of a trial usually remain the same through the creation of treatment arms by randomization.

Gerber et al. (2009) provide a study of the media's influence on political behavior. Media influence has been an important debate in the study of political behavior and is the subject of many studies using different methods, often with inconclusive results. A field experiment is able to provide a clear estimate of the impact of the media. In this study, carried out before the 2005 Virginia gubernatorial election in a northeastern Virginia county, Prince William, Gerber and colleagues randomly allocated free subscriptions of the more left-leaning paper the *Washington Post* and the more right-wing *Washington Times*. The county was selected because it was in the circulation area of both newspapers. The sample was selected by identifying through a survey 3,347 people who did not have a newspaper subscription. These people were randomized into 605 households to get the *Post*, 595 households to get the *Times*, and 904 households in the control condition. The researchers accessed state administrative records to measure the sample's voter turnout for the November 2005 and 2006 elections. Using a survey, carried out a week after the election, they found that those in the *Post* group were 8 percent more likely to vote Democrat than those in the control group, though they found a positive effect of voting Democrat for the *Times* even if at a less high rate than the *Post*. As the authors suggest, this result might be the impact of negative media reporting generally: getting news as a result of reading a newspaper from any standpoint rather than a particular newspaper's viewpoint influencing vote intentions toward the Democrat Party. Common in experiments, they found some null effects: there was "no effect of receiving either paper on knowledge of political events, opinions of those events, or on voter turnout in the 2005 gubernatorial election" (Gerber et al. 2009: 37).

It is also instructive to see how the researchers are careful about how much they extrapolate from the findings: how they deal with the question of external validity. The sample was a special population of people without newspaper subscriptions, so the findings do not apply to those who already subscribe to a paper. Nonetheless, because of randomization it is possible to make the inference that for this group of people the difference in outcomes is only attributable to getting the different newspapers because no other explanation is possible for the difference in voting outcomes between the two groups. Randomization is the key to making this inference.

In terms of thinking whether the assumptions behind the trial are breached, it is possible that members of the groups could have influenced each other, perhaps if they live next door and discuss what they read. In this case, it is unlikely as they are unlikely to be able to read each other's newspapers. Attrition can happen in the measurement of outcomes as people may not reply to the survey (there was a 32 percent response rate) or move house (unlikely in such a short period of time in the study). Moreover, 6 percent of households in the treatment groups opted out of the free subscription. Some addresses were not deliverable, which is another common experience in field experiments, and there were some households who had to be removed because they were already subscribing. These are the normal practical features of a field experiment that need to checked up and reported, but do not affect the internal validity of the research, as they reduced the sample size but did not introduce bias.

The practical demands of field trials

Trials as organizational tasks

The advantage of doing field experiments – providing a causal inference – must be balanced against their costs. Getting a field experiment off the ground is a complicated operation involving considerable ingenuity. In most social science research projects, attention is devoted to the collection of data, locating interviewees or data points, addressing ethical issues over the transfer of data or in dealing with people, and then ensuring that the data are analyzed correctly, such as with an appropriate statistical model, all of which is complex and time-consuming. All these considerations apply to a field experiment, but the researcher also needs to deliver an intervention or ensure that an intervention is delivered by someone else and takes place in a way that respects the integrity of the trial and its assumptions. As well as measuring behavior and attitudes, researchers need to manage an intervention or treatment, and then plan and monitor what happens to the participants, such as whether they drop out or make contact across the treatment and control groups. These organizational tasks require careful attention to detail. The units need to be randomized, and data collected from them as they proceed from the start of the experiment to the finish. There are many

practical tasks to be planned, such as the delivery of the treatment, and there will be several assessments of the best ways to harness the researcher's and organization's energy and resources. It is important to know in advance what the likely costs of doing an experiment in terms of time and energy expended are, as well as the financial resources to be consumed.

Implementing field experiments

Experiments risk failure or compromised implementation. The potential for error often arises by having to adapt plans to fast-changing situations. Unexpected snags crop up and Solomon-like decisions have to be made, sometimes with a deadline of a few hours. A field experiment relies on everything being in place at the right time and often depends on the cooperation of third parties. The implementation of a field experiment is a complex and interlocking project that needs careful management and requires an understanding of how all the different elements fit together. To make such a venture work, the researcher needs many skills beyond an understanding of their subject, such as the ability to persuade partners and considerable determination. In this book, there are plenty of examples that show that these practical considerations are important to think about, and most researchers spend as much time on them as all other activities put together, including project design, literature review, analysis, and writing up. If these things are not attended to, the experiment will be delivered poorly or not at all. All the time spent and resources dispersed would then be wasted.

Experiments need the smooth interrelationship of different processes that work to time and in the right order. Even when an experiment is implemented well in the sense that randomization took place, the interventions were administered properly and outcomes were measured effectively, there may be aspects of the implementation that undermine or limit the experiment – for example, some people from the control group inadvertently may have received the intervention but the senior management or researchers did not notice. The experiment looks like it worked out fine but, in fact, contains a hidden implementation failure. Because experiments are thought to deliver the gold standard and are widely believed to be superior over other research methods, policy-makers and other academics might assume the findings from such a faulty experiment are true and may start implementing the policy based on these results. This kind of implementation wastes resources and imposes costs on the recipients of the policy or intervention, as well as damages the reputation of experiments. Because of the possibility of implementation failure, the seasoned investigator worries about the experiment all the time and is continually on the phone, emailing and video-conferencing to ensure the project is on track. The researcher continually believes that she or he has not thought of something and their experiment is about to fail or will produce invalid results. The experimenter

thinks that if something can go wrong, it will! Or, they believe they will be hit by something totally unexpected.

The collaborative context

The importance of partnership

The reason why field experiments are often more constrained by practical considerations than other methods is that the researcher frequently relies on someone else to deliver the intervention. The delivery of the intervention can be done by team of workers, but more than likely is going to depend upon an organization, such as an association, group, or a public body, to do it. For much of political science and certainly for public policy, public decision-makers or policy-makers hold the key to doing experiments, because they can vary things on the researcher's behalf or want to test their own policies and need researchers to help them. Political science covers the procedures, institutions, and behaviors that allocate resources and make authoritative decisions in a society, such as voting for incumbents, legislators turning up to vote, and so on. Public policy is about any public decision that affects the outcomes citizens experience and can be about the means to ensure the effective administration of society (sometimes called public administration). It can include topics such as the efficiency of healthcare services, the extent to which governments can implement environmental policies, how governments run a cost-effective tax collection system, or the best ways to help people back into employment. In both political science and public policy, it can be important to know what happens when those public decisions vary, such as different campaign strategies or new ways to get people into employment, which is different from just observing them – hence the attraction of field experiments. In many of these cases, only policy-makers can deliver the real-world changes needed to appraise the research question. These trials can be called policy experiments. For political scientists, the equivalent to the policy-makers are the politicians, members of the interest group or political party workers, who can alter real world conditions on behalf of the researchers, either by themselves or through the organizations in which they are members.

The demands and needs the policy-makers

In spite of all the opportunities policy-makers offer, they also impose constraints that limit experiments. They can do this because researchers need their cooperation to get a project done. This veto is formal in the sense that experiments only takes place because a policy-maker has agreed to the project. However, the level of commitment is also important informally, including the motivation of senior and junior staff to implement the experiment. The initial agreement has to do more than just passively give

permission for the study to take place; the agency should actively support the project throughout and ensure that it happens. The experiment may fail if the employees of the organization do not fully buy into the project and follow the outlined procedures of a precise design. Experiments cost the policy-maker time and money, which may have been budgeted for by an external grant, but often have not. Even when funds are allocated to the delivery of an experiment, established procedures need to be changed by inserting randomization into the administrative process, ensuring the smooth, legal, and accurate transfer of data to third parties, briefing staff to deliver the intervention, and confirming they follow the right protocol and do not subvert the experiment or weaken it. Policy-makers have views about how every aspect of an experiment should be run: which population the sample should be drawn from; how many participants should be selected; how the randomization should be carried out; what should be included in the treatment; what data can and cannot be collected; whether there are ethical concerns concerning the people who get the intervention; how to match the data to the outcomes; how to deal with attrition and follow-up; and what can be reported and when. Policy-makers have costs to consider and may wish to minimize disruption caused by the field experiment even where there is a strong belief in its implementation. Policy-makers must put a lot of effort into ensuring the intervention takes place, as a lack of commitment at the top will be conveyed to staff down the line, and this lack of enthusiasm may lead to a weakly implemented experiment.

Political concerns may arise as to what findings the experiment will produce and whether they might challenge policy-makers. Agency staff and politicians may worry that randomization denies the treatment to clients that might depend on the public authority. A decision to allocate by the die may be negatively conveyed in the local media. Policy-makers might start off enthusiastic about experiments, but over time they may become less engaged as the initial warm glow about doing something innovative wears off and the perceptions of the costs and inconvenience come to the fore. Even though policy-makers may be enthusiastic about field experiments, they are primarily managers and report to superiors or wider constituencies; they have to give consideration to a wide range of concerns; and evaluation is not necessarily the most important issue they have to face on a day-to-day basis. The role of policy-makers is to balance out different considerations, not necessarily to follow every demand and need of the researchers. In the end, the researcher is in a weak position because the policy-maker may abandon a field experiment without much lost. Meanwhile, the cessation of a field experiment is very costly for researchers, equating to the end or a blow to the project and its funder and lost opportunities for publication that will harm the career of researchers.

There is considerable discretion from policy-makers about how to deliver the field experiment, and a lot of complex stages need to be worked through in a series of meetings. It is very unlikely that a simple trial design is

implemented just as the researcher wanted it in a protocol. Even if meetings agreeing to the trial design have taken place, new issues come up and snags are encountered. Moreover, the researcher is often an outsider and may not understand the policy environment or how the organization operates, which often is highly complex and involves technical issues, the use of information technology, personnel management, legal constraints, and contracts with external agencies – all of which are unfamiliar to someone who spends their time in an academic department or research institute. It is important to observe the delivery of the experiment to see how the rules and norms of the organization determine how a treatment is delivered – even to find out whether it is delivered at all.

Learning from policy-makers

The outsider researcher needs to learn from policy-makers or else the field experiment will fail. At each stage of the trial there are negotiations – whether it is the idea of doing the trial in the first place, agreeing to the exact design, deciding on the sample definition and size, determining the best means of randomization, agreeing on the use of data, or outlining the write up of the results data. Policy-makers will need to be educated into the arcane world of trials. Often they might not understand what randomization entails or that it is important to sometimes deny services to some clients for a short period of time. The researcher needs to persuade policy-makers to take note of these things, or there is a danger that lack of knowledge will determine the fate of the trial and errors or weaknesses will ensue. Likewise, there is often not just one policy-maker but several, reflecting different parts of the organization, such as the senior policy-makers, those responsible for delivery, and those in separate units responsible for data collection and information technology management. There may be several organizations that need to work together, such as the donor organization, the recipient government, regional or local governments, and the delivery organization, as well as voluntary groups.

There are many benefits to working with policy-makers as well as constraints. It is often the case that policy-makers are able to save a trial from failure by thinking of a more practical way to deliver it. The practical skills of policy-makers can help to overcome the logistical challenges of delivery. The partnership is two-way with researchers thinking like policy-makers and policy-makers thinking like researchers for mutual benefit. There may be hidden opportunities for randomization that emerge from discussions and meetings between the two.

Varieties of policy experiments

Policy experiments differ from each other. Policy-makers sometimes commission experiments, so they want them to work well. A powerful external

funder can create the incentive for the policy-maker to do the experiment. Some of these official policy experiments are very well funded and can build in procedures to ensure the experiment is conducted to a very high standard. The funder can determine whether researchers can carry out aspects of the intervention or data collection, making them less reliant on the policy-maker or allowing them freedom to check that the intervention is being carried out properly. Others may be based on a partnership, which might come from the initiative of the researchers or because a particular manager has an idea and good contacts with the researcher. When there are fewer resources, an experiment is done quickly and without a dedicated team, what might be called informal policy experiments or opportunistic experiments. Opportunities for interventions arise that are quite cheap to do. Many organizations are discovering that they can randomize easily and with little cost. It is easy to vary randomly standard procedures, reducing the costs of the extra treatment to research time (see Rogers 2014). Many organizations have their own evaluators and who collect and analyze the data. Given the need for expertise, they might commission external organizations to evaluate for them, but only pay for days needed to design the experiment and do the analysis; everything else except for extra meetings can be internalized, so even commissioning is cheap compared to the costs of the intervention itself. Academics are very useful for these kinds of trials, as they may encourage the organization to do one and then help with the design and analysis, as it gives them access to an experiment they would not otherwise be able to do. However, given these experiments might happen more because of opportunity, opportunistic interventions face higher risks than longer-term evaluations. The risk occurs because resources might not be in place to monitor the intervention, and quick interventions might have flaws if designed on the fly or if there is no follow-through. Because no money changes hands and there is no formal evaluation, the policy-maker might not commit to such an experiment. It may be harder for the researcher to challenge policy-makers, say if she suspects the treatment has not been delivered properly, because making such an assertion breaches the trust and goodwill that created the collaboration in the first place. In fact, this weakness of the control of the design need not occur, and perhaps these informal experiments might have more value than long running evaluations. Nimbleness and enthusiasm can pay dividends, but there are risks, too.

Political science experiments

Many political science experiments do not always require the researcher to deal directly with policy-makers, but some of the same issues arise with these interventions. Experimentalists still have to grapple with a lot of practical issues that emerge from the field context and the existence of other organizations, such as whether public authorities need to be notified, whether the research is compliant with the law, whether other interventions are

happening at the same time, how to create a field-force that will implement the experiment fully, and how to respond to the snags that come up and have to be solved on a daily basis. As will be shown in Chapter 4, political science started with experiments where they could organize mobilization without policy-makers and other stakeholders. But as experiments have grown more complicated, political scientists have found policy-makers or other people in public life are needed to carry out interventions on their behalf, whether they are political parties with partisan experiments, interest groups for lobbying experiments, or politicians or bureaucrats for elite experiments.

If political scientists would rather deliver an intervention themselves, they might run the risk of pretending to be the policy-maker or advocate, which then creates legal and ethical issues about misrepresentation and taking on a false identity. A controversy occurred when, in October 2014, researchers at Stanford University and the University of Dartmouth sent 100,000 flyers to Montana residents offering opinions about candidates for the Supreme Court elections. They used the State of Montana's official seal without consulting the state, effectively impersonating a policy-maker by using the emblem on the flyer (Willis 2014). Had the policy-maker agreed to the intervention, then no such controversy would have arisen; but then the researchers would have had to negotiate with the policy-maker to use the state's seal on their flyer, and these policy-makers would have worked to ensure their preferences were met in the research design before agreeing to it, pushing it back in the realm of the policy experiments discussed above.

Plan of the book

Summary of the argument

This chapter has introduced field experiments by explaining what they are and why they reveal causal inference. It has sought to familiarize the reader with the different aspects of a randomized experiment to convey something about the complexity of carrying one out. Above all, it has stressed that an experiment requires an intervention to be delivered alongside the more familiar processes of measurement and data analysis. The intervention needs to be fully integrated with the procedures for measurement. In some cases, researchers carry out the intervention themselves by, for example, mailing or telephoning voters; in other cases, in what are called policy experiments, this burden can be borne by personnel located in an agency whose staff members implement the intervention and collect much of the data. The latter approach leverages the power of policy-makers to change things of importance in politics, and the potential to change outcomes is very great – probably more so than researchers acting on their own; but the cost of this approach is the need to negotiate the design of the trial with policy-makers who may limit what the researcher wants to do (though practitioners usually help the researcher, too). These constraints, particularly those limiting

where and on whom the trial takes place, affect the extent to which the researchers can generalize from a trial. In addition, the delegation of responsibilities for the intervention and measuring outcomes create the potential for implementation slippage, threatening the integrity of an experiment especially where monitoring is hard to do. There is a risk that failing to implement a trial fully or practical obstacles can limit the inferences that can be made from a trial or even invalidate it completely. Every trial in public policy and political science risks failure of implementation, and even small weaknesses can undermine an experiment. This book is an exploration of these practical constraints on trials and a discussion of the best ways to overcome them. The key message of this book is that it is possible to carry out a successful trial with enough planning, forethought and vigilance: the watchful researcher can implement a successful experiment. The challenge is to build up practical knowledge about trial design, use common sense to work out what needs to be done during the trial, and react in a sensible way when crises arise.

General overview and background

Chapter 2 walks the reader through the ten steps needed to deliver a trial, examining the tasks that need to be considered when designing it. These steps are related and must interlock when planning the experiment. Chapter 3 deals with the practical issues that arise when implementing a trial and how it often has to be modified as new information comes to light; this consideration places further constraints on the experiments though it can also provide extra opportunities. The chapter lists out and explains the nine most common threats to experiments. Chapter 4 presents the whole picture, tracing the development of field trials in political science and public policy from the 1920s until now to give the context and direction of travel, as well as offering a reflection on why experiments started in the fields they did and grew in popularity in certain areas. It tries to account for why their development in political science and public policy has been relatively slow, at first at least. The chapter considers the argument that practical opportunities generated the areas for expansion and that real and perceived costs have limited the use of trials.

The case studies

The remainder of the book is a series of case studies of areas that have attracted attention, particularly in recent years. The idea is to explore in more depth how political scientists and policy scholars have carried out trials and used them to advance knowledge, using the ten steps of design and the nine common threats as guides. Each chapter has the same structure: there is a brief introduction followed by an account of how a trial is carried out in a particular field. After a short history of how trials developed in this area, the core part

of each chapter uses examples from prominent studies to showcase each of the ten design steps in turn. The following section has a similar approach by taking on each of the nine threats to experiments and in using examples.

The range of topics in political science and public policy is potentially vast, and the book focuses on those topics that have received particular attention from experimentalists. It is not possible to cover all areas of public policy because that would be too big a task. There is no chapter on health-related interventions, nor one directly on education or crime. In spite of these omissions, the hope is that the book covers enough of a variety of the kinds of trials conducted to highlight the challenges and opportunities that exist as well as compare the various fields of enquiry.

The sources of information for the case studies in this book are mainly published studies and grey sources, such as official reports, that have been read carefully for details on research design. Great care has been taken to review the design sections, footnotes, appendices, and annexes, which are often where information about the design and implementation of a field experiment is contained. The idea is to give extensive detail about design choices, and each chapter gives examples of problems researchers have faced and solved. Many textbooks and guides tend not to give that many examples (e.g., Solomon et al. 2009; Hutchinson and Styles 2010), nor they do not offer a large amount of detail from published studies (e.g., Glennerster and Takavarasha 2013). Others focus on reviewing different aspects of the method (Torgerson and Torgerson 2008) or explaining the causal reasoning and the implementation of statistical tests (Morton and Williams 2010; Gerber and Green 2012). In contrast, this book gives many examples where it is possible to see the link between theory, method, and implementation. It should be seen as a companion to the other texts, especially Gerber and Green (2012) and Glennerster and Takavarasha (2013).

To find out more about how the trials worked in practice, interviews with experimentalists were carried out. To respect the interviewees, this information has mainly been used for background and to help understand the implementation of studies. The observations made in the text are in the main related to published studies. The interviewees were helpful in drawing attention to particular phases of the implementation process in their studies. The exception is where the author is talking about his own work, but even here there has been a need to protect partners in agencies so that many of these details have been elided too. Of course, the information used in this book is not experimental, nor is it based on a wide-ranging survey of researchers. It relies on the reader agreeing to or at least being provoked by the reviews of research practice. Essentially, it is a guide and relies on the reader finding the review and detail supplied to be plausible.

The case studies are as follows. Chapter 5 looks at a core area in political science: the mobilization of voters to turn out at the polls or act in other ways that are political, such as joining an interest group. Chapter 6 examines experiments on elites, which is a more controversial policy area in political

science that shows some of the limits of trials and their ethical constraints. Chapter 7 takes on the field of welfare and employment, in particular welfare reform, which has been an important example of policy experiments carried out over many decades. Chapter 8 is on experiments to nudge citizens – behavioral experiments that have become popular in recent years. Chapter 9 concerns experiments in developing countries using examples that are either political science or policy interventions and are carried out in very different contexts than those in settled developed world contexts. Chapter 10 pulls it all together, seeking to draw some common lessons from these diverse trials.

References

Adams, William C. and Dennis J. Smith. 1980. "Effects of Telephone Canvassing on Turnout and Preferences: A Field Experiment." *Public Opinion Quarterly* 44: 389–395.

Aiken, Alexander M., Calum Davey, James R. Hargreaves, and Richard J. Hayes. 2015. "Re-analysis of Health and Educational Impacts of a School-based Deworming Programme in Western Kenya: A Pure Replication." *International Journal of Epidemiology* Early view.

Ainsworth, Hannah R., David J. Torgerson, and Arthur R. Kang 'Ombe. 2010. "Conceptual, Design, and Statistical Complications Associated With Participant Preference." *The ANNALS of the American Academy of Political and Social Science* 628: 176–188.

Andersen, Lotte Bøgh, Niels Westergaard Nielsen, Anne Bøllingtoft, Jacob Ladenburg, Christian Bøtcher Jacobsen, Poul Aaes Nielsen, Tine Eriksen, Ann-Louise Holten, Ulrich Thy Jensen, and Louise Ladegaard Bro. 2014. "Leadership Training, Leadership Strategies and Organizational Performance: Designing An Experiment to Test the Causal Effect of Leadership on Performance." Paper to workshop Causal Effects in Political Science: The Promise and Pitfalls of Experimental Methods at the Nordic Political Science Association Conference in Gothenburg, August 12–15, 2014.

Andersen, Simon Calmar and Donald P. Moynihan. 2014. "Bureaucratic Investments in Expertise: Evidence From a Randomized Controlled Field Trial." Paper prepared for presentation at the 72nd annual MPSA Conference, Chicago.

Angrist, Joshua D., Eric Bettinger, Erik Bloom, Elizabeth King, and Michael Kremer. 2002. "Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment." *The American Economic Review* 92(5): 1535–1558.

Angrist, Joshua D., Eric Bettinger and Michael Kremer. 2006. "Long-Term Educational Consequences of Secondary School Vouchers: Evidence from Administrative Records in Colombia." *American Economic Review* 96: 847–862.

Arceneaux, Kevin. 2007. "I'm Asking for Your Support: The Effects of Personally Delivered Campaign Messages on Voting Decisions and Opinion Formation." *Quarterly Journal of Political Science* 1: 43–65.

Arceneaux, Kevin and Daniel M. Butler. 2016. "How Not to Increase Participation in Local Government: The Advantages of Experiments When Testing Policy Interventions." *Public Administration Review* 76: 131–139.

Ariely, Dan, Anat Bracha, and Stephan Meier. 2009. "Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially." *American Economic Review* 99: 544–555.

Avellaneda, Claudia N. 2013. "Mayoral Decision-making: Issue Salience, Decision Context, and Choice Constraint? An Experimental Study With 120 Latin American Mayors." *Journal of Public Administration Research and Theory* 23: 631–661.

Banerjee, Abhijit V. and Esther Duflo. 2014. "The Experimental Approach to Development Economics." In *Field Experiments and Their Critics: The Uses and Abuses of Experimentation in the Social Sciences*, edited by Dawn Langan Teele, 78–114. New Haven, CT: Yale University Press.

Banerjee, Abhijit, Dean Karlan, and Jonathan Zinman. 2015. "Six Randomized Evaluations of Microcredit: Introduction and Further Steps." *American Economic Journal: Applied Economics* 7: 1–21.

Barnow, Burt S. 2010. "Setting Up Social Experiments: The Good, the Bad, and the Ugly." *Zeitschrift für Arbeitsmarkt Forschung* 43: 91–105.

Baron, Jon. 2013. *Randomized Controlled Trials Commissioned by the Institute of Education Sciences Since 2002: How Many Found Positive Versus Weak or No Effects*. http://coalition4evidence.org/wp-content/uploads/2013/06/IES-Commissioned-RCTs-positive-vs-weak-or-null-findings-7-2013.pdf. Accessed July 30, 2015.

Barrett, Christopher B. and Michael R. Carter. 2010. "The Power and Pitfalls of Experiments in Development Economics: Some Non-random Reflections." *Applied Economic Perspectives and Policy* 32: 515–548.

Barrett, Christopher B. and Michael R. Carter. 2014. "Retreat from Radical Skepticism: Rebalancing Theory, Observational Data, and Randomization in Development Economics." In *Field Experiments and Their Critics: The Uses and Abuses of Experimentation in the Social Sciences*, edited by Dawn Langan Teele, 58–77. New Haven, CT: Yale University Press.

Beath, Andrew, Christia Fontina, and Ruben Enikolopov. 2013. "Empowering Women through Development Aid: Evidence From a Field Experiment in Afghanistan." *American Political Science Review* 107: 540–557.

Bedolla, Lisa García and Melissa R. Michaelson. 2012. *Mobilizing Inclusion: Transforming the Electorate through Get-Out-the-Vote Campaigns*. New Haven, CT: Yale University Press.

Behavioural Insights Team. 2012. *Fraud, Error and Debt: Behavioural Insights Team Paper*. London: Cabinet Office.

Behavioural Insights Team, Department of Health, Driver and Vehicle Licensing Agency and NHS Blood and Transplant. 2013. *Organ Donor Registrations: Trialling Different Approaches*. London: BIT. https://www.gov.uk/government/publications/organ-donor-registrations-trialling-different-approaches. Accessed August 4, 2015.

Behavioural Insights Team. 2015. *The Behavioural Insights Team Update 2013–2015*. London: Behavioural Insights Team.

Belot, Michèle and Jonathan James. 2013. "Partner Selection into Policy Relevant Field Experiments" (No. 236). Edinburgh School of Economics, University of Edinburgh.

Bergan, Daniel. E. 2009. "Does Grassroots Lobbying Work? A Field Experiment Measuring the Effects of an E-mail Lobbying Campaign on Legislative Behavior." *American Politics Research* 37: 327–352.

Berk, Richard A., Gordon K. Smyth, and Lawrence W. Sherman. 1988. "When Random Assignment Fails: Some Lessons from the Minneapolis Spouse Abuse Experiment." *Journal of Quantitative Criminology* 4: 209–224.

Bertrand, Marianne, Simeon Djankov, Rema Hanna, and Sendhil Mullainathan. 2007. "Obtaining a Driver's License in India: An Experimental Approach to Studying Corruption." *The Quarterly Journal of Economics* 122: 1639–1676.

Blattman, Christopher, Julian C. Jamison, and Margaret Sheridan. 2015. "Reducing Crime and Violence: Experimental Evidence on Adult Noncognitive Investments in Liberia." Unpublished paper. Available at SSRN: http://ssrn.com/abstract=2594868 or http://dx.doi.org/10.2139/ssrn.2594868, Accessed August 4, 2015.

Bloom, Howard S. 2004. "Randomizing Groups to Evaluate Place-based Outcomes." In *Learning More from Social Experiments: Evolving Analytic Approaches*, edited by Howard S. Bloom, 115–172. New York: Russell Sage.

Bloom, Howard S., Carolyn J. Hill, and James A Riccio. 2003. "Linking Program Implementation and Effectiveness: Lessons from a Pooled Sample of Welfare-to-Work Experiments." *Journal of Policy Analysis and Management* 22: 551–575.

Bloom, Howard S., Larry L. Orr, Stephen H. Bell, George Cave, Fred Doolittle, Winston Lin, and Johannes M. Bos. 1997. "The Benefits and Costs of JTPA Title II-A Programs: Key Findings from the National Job Training Partnership Act Study." Paper prepared for presentation at the 72nd annual MPSA Conference, Chicago.

Blume, Toby and Peter John. 2014. *Using Nudges to Increase Council Tax Collection: Testing the Effects through a Randomized Controlled Trial*. London: Lambeth Borough Council.

Blustein, Jan. 2005. "Toward a More Public Discussion of the Ethics of Federal Social Program Evaluation." *Journal of Policy Analysis and Management* 24: 824–846.

Boggis, J. G. 1974. *Domestic Tariffs Experiment*. Load and Market Research Report no. 121. London: The Electricity Council.

Bond, Robert M., Christopher J. Fariss, Jason J. Jones, Adam D. I. Kramer, Cameron Marlow, Jaime E. Settle, and James H. Fowler. 2012. "A 61-Million-Person Experiment in Social Influence and Political Mobilization." *Nature* 489: 295–298.

Bowers, Jake. 2011. "Making Effects Manifest in Randomized Experiments." In *Cambridge Handbook of Experimental Political Science*, edited by James N. Druckman, Donald P. Green, James H. Kuklinski, and Arthur Lupia, 459–480. Cambridge: Cambridge University Press.

Bowers, Jake, Mark M. Fredrickson, and Costas Panagopoulos. 2013. "Reasoning About Interference Between Units: A General Framework." *Political Analysis* 21(1): 97–124.

Box, Joan F. 1980. "R. A. Fisher and the Design of Experiments, 1922–1926." *The American Statistician* 34: 1–7.

Bro, Louise Ladegaard, Lotte Bøgh Andersen, Anne Bøllingtoft, Tine Eriksen, Ann-Louise Holten, Christian Bøtcher Jacobsen, Ulrich Thy Jensen, Jacob Ladenburg, Poul Aaes Nielsen & Niels Westergaard-Nielsen. 2015. "Leadership Training, Leadership Strategies and Organizational Performance. Designing an Experiment to Test the Causal Effect of Leadership on Performance." Unpublished paper. Available at http://ps.au.dk/en/research/research-projects/leap-leadership-and-performance/publications-and-papers/. Accessed September 1, 2015.

Broockman, David. E. and Daniel M. Butler. 2014. "How Do Citizens React When Politicians Support Policies They Oppose? Field Experiments with Elite

Communication." Unpublished paper. http://stanford.edu/~dbroock/papers/broockman_butler_legislator_experiments_2014.pdf. Accessed July 31, 2015.

Burghardt, John, Peter Z. Schochet, Sheena McConnell, Terry Johnson, R. Mark Gritz, Steven Glazerman, John Homrighausen, and Russell Jackson. 2001. *Does Job Corps Work: Summary of the National Job Corps Study*. Princeton, NJ: Mathematica Policy Research, Inc. (MPR Ref. 8140–8530).

Butler, Daniel M. 2010. "Monitoring Bureaucratic Compliance: Using Field Experiments to Improve Governance." *Public Sector Digest* 2010(winter): 41–44.

Butler, Daniel M. 2014. *Representing the Advantaged: How Politicians Reinforce Inequality*. Cambridge: Cambridge University Press.

Butler, Daniel M. and David E. Broockman. 2011. "Do Politicians Racially Discriminate Against Constituents? A Field Experiment on State Legislators." *American Journal of Political Science* 55: 463–477.

Butler, Daniel M., Christopher Karpowitz, and Jeremy C. Pope. 2012. "A Field Experiment on Legislators' Home Style: Service versus Policy." *Journal of Politics* 74(2): 474–486.

Butler, Daniel M. and David W. Nickerson. 2011. "Can Learning Constituency Opinion Affect How Legislators Vote? Results from a Field Experiment." *Quarterly Journal of Political Science* 6: 55–83.

Cabinet Office and Charitable Aid Foundation. 2013. *Applying Behavioural Insights to Charitable Giving*. London: Cabinet Office. www.gov.uk/government/uploads/system/uploads/attachment_data/file/203286/BIT_Charitable_Giving_Paper.pdf. Accessed August 2, 2015.

Cabinet Office and Institute for Government. 2010. *MINDSPACE: Influencing Behaviour through Public Policy*. London: Cabinet Office/Institute for Government.

Campbell, Donald T. 1957. "Factors Relevant to the Validity of Experiments in Social Settings." *Psychological Bulletin* 54: 297–312.

Campbell, Donald T. 1969. "Reforms as Experiments." *American Psychologist* 24: 409–429.

Campbell, Donald T. and Julian Stanley. 1963. *Experimental and Quasi-Experimental Designs for Research*. Dallas: Houghton Mifflin, 1st edition.

Chalmers, Iain. 2007. "Joseph Asbury Bell and the Birth of Randomized Trials." *Journal of the Royal Society of Medicine* 100: 287–293.

Chande, Raj, Michael Luca, Michael Sanders, Xia Zhi Soon, Oana Borcan, Netta Barak-Corren, Elizabeth Linos, Elspeth Kirkman, and Sean Robinson. 2015. "Curbing Adult Student Attrition: Evidence from a Field Experiment." *Harvard Business School Working Paper* No. 15–065.

Chattopadhyay, Raghabendra and Esther Duflo. 2004. "Women as Policy Makers: Evidence from a Randomized Policy Experiment in India." *Econometrica* 72: 1409–1443.

Chin, Michele L., John R. Bond, and Nehemia Geva. 2000. "A Foot in the Door: An Experimental Study of PAC and Constituency Effects on Access." *The Journal of Politics* 62: 534–549.

Chong, Alberto, Ana L. De La O, Dean Karlan and Leonard Wantchekon. 2015. "Does Corruption Information Inspire the Fight or Quash the Hope? A Field Experiment in Mexico on Voter Turnout, Choice, and Party Identification." *The Journal of Politics* 77: 55–71.

Cialdini, Robert B., Raymond R. Reno, and Carl A. Kallgren. 1990. "A Focus Theory of Normative Conduct: Recycling the Concept of Norms to Reduce

Littering in Public Places." *Journal of Personality and Social Psychology* 58: 1015–1026.

Citrin, Jack, Donald P. Green, and Levy Morris. 2014. *Election Law Journal: Rules, Politics, and Policy* 13(2): 228–242.

Cochrane, Archie L. 1972. *Effectiveness and Efficiency: Random Reflections on Health Services*. London: The Nuffield Provincial Hospitals Trust.

Cochrane, Archie L. and Max Blythe. 1989. *One Man's Medicine: An Autobiography of Professor Archie Cochrane*. Cardiff: University of Cardiff.

Conroy-Krutz, Jeffrey and Devra C. Moehler. 2016. "Partisan Media and Engagement: A Field Experiment in a Newly Liberalized System." *Political Communication* 33: 414–432.

Coppock, Alexander. 2014. "Information Spillovers: Another Look at Experimental Estimates of Legislator Responsiveness." *Journal of Experimental Political Science* 1(2): 159–169.

Coppock, Alexander. 2015. *Package 'randomizr' Easy to Use Tools for Common Forms of Random Assignment*, Version 0.3.0. https://alexandercoppock.com/randomizr/. Accessed October 23, 2016.

Dahl, Robert A. 1961. *Who Governs? Democracy and Power in an American City*. New Haven: Yale University Press.

Daly, Jeanne. 2005. *Evidence-Based Medicine and the Search for a Science of Clinical Care*. Berkley and London: University of California Press.

Davey, Calum, Alexander M. Aiken, Richard J. Hayes, and James R. Hargreaves. 2015. "Re-analysis of Health and Educational Impacts of a School-based Deworming Programme in Western Kenya: A Statistical Replication of a Cluster Quasi-randomized Stepped-Wedge Trial." *International Journal of Epidemiology* DOI: 10.1093/ije/dyv128.

Deaton, Angus. 2009. "Instruments of Development: Randomization in the Tropics, and the Search for the Elusive Keys to Economic Development." *Proceedings of the British Academy, 2008 Lectures* 162: 123–160.

Decker, Paul T., Robert B. Olsen, Lance Freeman, and Daniel H. Klepinger. 2000. *Assisting Unemployment Insurance Claimants: The Long-Term Impacts of the Job Search Assistance Demonstration*. Report to U.S. Department of Labor. Washington, DC.

Decker, Paul T. and Craig V. Thornton. 1995. "The Long Term Effects of Transitional Employment Services." *Social Security Bulletin* 58(4): 71–81.

Dennis, Michael A. 1990. "Assessing the Validity of Randomized Field Experiments: An Example from Drug Abuse Treatment Research." *Evaluation Review* 14: 347–373.

Dennis, Michael A. 1994. "Ethical and Practical Randomized Field Experiments." In *Handbook of Practical Program Evaluation*, edited by Joseph S. Wholey, Harry P. Hatry, and Kathryn E. Newcomer. San Francisco, CA: Jossey-Bass.

Department for Business and Skills and Cabinet Office. 2014. *Growth Vouchers Trial Protocol*. London: Department for Business and Skills and Cabinet Office. www.behaviouralinsights.co.uk/wp-content/uploads/2015/07/bis-14-561-growth-vouchers-programme-trial-protocol.pdf. Accessed August 2, 2015.

Desposato, Scott. 2017. Ed. *Ethics and Experiments: Problems and Solutions for Social Scientists and Policy Professionals*. New York: Routledge/Taylor & Francis.

Devine, Joel, Charles J. Brody, and James D. Wright. 1997. "Evaluating an Alcohol and Drug Treatment Program for the Homeless: An Econometric Approach." *Evaluation and Program Planning* 20: 205–215.

De Bondt, Werner F. M. and Richard Thaler. 1985. "Does the Stock Market Overreact?" *The Journal of Finance* 40: 793–805.

De La O, Ana and Fernando Martel. 2014. "Do Federal and State Audits Increase Compliance with a Grant Program to Improve Municipal Iinfrastructure: Atudy Protocol for a Randomized Controlled Trial." *BMC Public Health* 14: 912.

De La O, Ana and Leonard Wantchekon. 2011. "Experimental Research on Democracy and Development." In *Cambridge Handbook of Experimental Political Science*, edited by James N. Druckman, Donald P. Green, James H. Kuklinski, and Arthur Lupia, 384–396. Cambridge: Cambridge University Press.

De Vries, Catherine, Elias Dinas, and Hector Solaz. 2015. "You Have Got Mail! A Field Experiment on Legislator Responsiveness in the European Parliament." Paper to the Southern Political Science Association meeting, New Orleans, January 15–17.

Distelhorst, Greg and Yue Hou. 2014. "Ingroup Bias in Official Behavior: A National Field Experiment in China." *Quarterly Journal of Political Science* 9: 203–230.

Druckman, Jamie N., Donald P. Green, James Kuklinski, and Arthur Lupia, eds. 2011. *Cambridge Handbook of Experimental Political Science*. Cambridge: Cambridge University Press.

Druckman, Jamie N. and Arthur Lupia. 2006. "The Growth and Development of Experimental Research Political Science." *American Political Science Review* 100: 627–636.

Dubner, Stephen J. and Steven D. Levitt. 2005. *Freakonomics: A Rogue Economist Explores the Hidden Side of Everything*. New York: William Morrow/ HarperCollins.

Duflo, Esther. 2006 "Field Experiments in Development Economics." Discussion paper. http://economics.mit.edu/files/800. Accessed August 3, 2015.

Duflo, Esther and Rema Hanna. 2006. "Monitoring Works: Getting Teachers to Come to School." *NBER Working Paper* No. 11880.

Duflo, Esther, Rema Hanna, and Stephen P. Ryan. 2012. "Incentives Work: Getting Teachers to Come to School." *American Economic Review* 102: 1241–1278.

Duflo, Esther, Rachel Glennerster, and Michael Kremer. 2006. "Using Randomization in Development Economics Research: A Toolkit." National Bureau of Economic Research. *NBER Technical Working Paper* No. 333.

Dunning, Thad. 2012. *Natural Experiments in the Social Sciences. A Design-Based Approach*. Cambridge: Cambridge University Press.

Einstein, Katherine Levine and David M. Glick. 2015. "Does Race Affect Access to Government Services? An Experiment Exploring Street Level Bureaucrats and Access to Public Housing." Unpublished paper. http://dornsife.usc.edu/assets/sites/741/docs/Panel_2a_paper4_EinsteinGlick_Does_Race_Affect_Access_to_Housing_SoCLASS_USC_2015.pdf. Accessed August 1, 2015.

Eldersveld, Samuel J. 1956. "Experimental Propaganda Techniques and Voting Behaviour." *The American Political Science Review* 50: 15–165.

Erikson, Robert S. and Laura Stoker. 2011. "Caught in the Draft: The Effects of Vietnam Draft Lottery Status on Political Attitudes." *American Political Science Review* 10: 221–237.

Ettelt, Stefanie, Nicholas Mays, and Pauline Allen. 2015. "Policy Experiments: Investigating Effectiveness or Confirming Direction?" *Evaluation* 21: 292–307.

Fang, Albert H., Andrew M. Guess, and Macartan Humphreys. 2014. *Messaging Strategies to Combat Housing Discrimination in New York City: Evidence from a Field*

Experiment. Evaluation Report. Draft 2.0. Prepared for the New York City Commission on Human Rights. New York: Columbia University.

Farrington, David and Doris Layton MacKenzie. 2013. "Long-term Follow-ups of Experimental Interventions." *Journal of Experimental Criminology* 9: 385–388.

Farrington, David P. and Brandon C. Welsh. 2005. "Randomized Experiments in Criminology: What Have We Learned in the Last Two Decades?" *Journal of Experimental Criminology* 1: 9–38.

Faulkner, William N. 2014. "A Critical Analysis of a Randomized Controlled Trial Evaluation in Mexico: Norm, Mistake Or Exemplar?" *Evaluation* 20: 230–243.

Fieldhouse, Ed, Dave Cutts, Paul Widdop, and Peter John. 2013. "Do Impersonal Mobilisation Methods Work? Evidence from a Nationwide Get-Out-The-Vote Experiment in England." *Electoral Studies* 32: 113–123.

Fieldhouse, Ed, Dave Cutts, Peter John, and Paul Widdop. 2014. "When Context Matters: Assessing Geographical Heterogeneity of Get-Out-The-Vote Treatment Effects Using a Population Based Field Experiment." *Political Behavior* 36: 77–97.

Fisher, Ronald A. 1926. "The Arrangement of Field Experiments." *Journal of the Ministry of Agriculture of Great Britain* 33: 503–513.

Fisher, Ronald A. 1935. *The Design of Experiments*. Edinburgh: Oliver and Boyd.

Foos, Florian and Peter John. 2015. "Parties Are No Civic Charities: Campaigns, Demobilisation, and the Changing Composition of the Electorate." Unpublished paper.

Forsetlund, Louise, Iain Chalmers, and Arild Bjørndal. 2007. "When Was Random Allocation First Used to Generate Comparison Groups in Experiments to Assess the Effects of Social Interventions?" *Economics of Innovation and New Technology* 16: 371–384.

Fujiwara, Thomas and Leonard Wantchekon. 2013. "Can Informed Public Deliberation Overcome Clientelism? Experimental Evidence from Benin." *American Economic Journal: Applied Economics* 5(4): 241–255.

Gerber, Alan S. and Donald P. Green. 2000a. "The Effects of Canvassing, Telephone Calls, and Direct Mail on Voter Turnout." *American Political Science Review* 94: 653–663.

Gerber, Alan S. and Donald P. Green. 2000b. "The Effect of a Nonpartisan Get-Out-The-Vote Drive: An Experimental Study of Leafletting." *Journal of Politics* 62 (3): 846–857.

Gerber, Alan S. and Donald P. Green. 2001. "Do Phone Calls Increase Voter Turnout? A Field Experiment." *Public Opinion Quarterly* 65: 75–85.

Gerber, Alan S. and Donald P. Green. 2005. "Correction to Gerber and Green (2000), Replication of Disputed Findings, and Reply to Imai (2005)." *American Political Science Review* 99(2): 301–313.

Gerber, Alan S. and Donald P. Green. 2012. *Field Experiments: Design, Analysis, and Interpretation*. New Haven, CT: Yale University Press.

Gerber, Alan S., Donald P. Green, Edward H. Kaplan, and Holger L. Kern. 2010. "Baseline, Placebo, and Treatment: Efficient Estimation for Three-Group Experiments." *Political Analysis* 18: 297–315.

Gerber, Alan S., Donald P. Green, and Christopher W. Larimer. 2008. "Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment." *American Political Science Review* 102: 33–48.

Gerber, Alan S., Dean Karlan, and Daniel Bergan. 2009. "Does the Media Matter? A Field Experiment Measuring the Effect of Newspapers on Voting Behavior and Political Opinions." *American Economic Journal: Applied Economics* 1: 35–52.

Gimpel, James, Daron Shaw, and Wendy Tam Cho. 2005. "Turning Out the Vote in Texas." Paper presented at the Annual Meeting of the American Political Science Association, Marriott Wardman Park, Omni Shoreham, Washington Hilton, Washington, DC, September 1.

Giné, Xavier and Ghazala Mansuri. 2011. "Together We Will: Experimental Evidence on Female Voting Behavior in Pakistan." *Policy Research Working Papers*, June 2011

GiveWell. 2012. "Reanalysis of the Miguel and Kremer Deworming Experiment." www.givewell.org/international/technical/programs/deworming/reanalysis#External validity. Accessed August 4, 2015.

Glennerster, Rachel and Kudzai Takavarasha. 2013. *Running Randomized Evaluations: A Practical Guide*. Princeton, NJ: Princeton University Press.

Glewwe, Paul, Albert Park, and Meng Zhao. 2014. "A Better Vision for Development: Eyeglasses and Academic Performance in Rural Primary Schools in China." Working Paper, June 2014. http://s3.amazonaws.com/fieldexperiments-papers/papers/00254.pdf. Accessed August 6, 2015.

Glewwe, Paul, Michael Kremer, and Sylvie Moulin. 2000. "Textbooks and Test Scores: Evidence from a Prospective Evaluation in Kenya." Policy Research Group. The World Bank. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.196.3625&rep=rep1&type=pdf. Accessed August 6, 2015.

Glewwe, Paul, Michael Kremer, Sylvie Moulin, and Eric Zitzewitz. 2004. "Retrospective vs. Prospective Analyses of School Inputs: The Case of Flip Charts in Kenya." *Journal of Development Economics* 74: 251–268.

Gonzalez-Ocantos, Ezequiel, Chad Kiewiet de Jonge, Carlos Meléndez, Javier Osorio, and David W. Nickerson. 2011. "Vote Buying and Social Desirability Bias: Experimental Evidence from Nicaragua." *American Journal of Political Science* 56: 202–217.

Gosnell, Harold F. 1926. "An Experiment in the Stimulation of Voting." *The American Political Science Review* 20: 869–874.

Gosnell, Harold F. 1927. *Getting-Out-The-Vote: An Experiment in the Stimulation of Voting*. Chicago, IL: Chicago University Press.

Green, Donald P. 2009. "Regression Adjustments to Experimental Data: Do David Freedman's Concerns Apply to Political Science?" Working paper. Society of Political Methodology. http://polmeth.wustl.edu/mediaDetail.php?docId=917. Accessed August 3, 2015.

Green, Donald P., Peter M. Aronow, and Mary C. McGrath. 2010. "Making Sense of 200+ Field Experiments on Voter Mobilization, Part I: Direct Mail." Paper prepared for presentation at the Annual Meeting of the American Political Science Association, Washington, DC, September 2–5.

Green, Donald P. and Alan S. Gerber. 2003. "The Underprovision of Experiments in Political Science." *The ANNALS of the American Academy of Political and Social Science* 589: 94–112.

Green, Donald P. and Alan S. Gerber. 2008. *Get Out the Vote: How to Increase Voter Turnout*. Washington, DC: Brookings Institution Press, 2nd edition.

Green, Donald P. and Alan S. Gerber. 2015. *Get Out the Vote: How to Increase Voter Turnout*. Washington, DC: Brookings Institution Press, 3rd edition.

Green, Donald P., Alan S. Gerber, and David W. Nickerson. 2003. "Getting Out the Vote in Local Elections: Results from Six Door-to-Door Canvassing Experiments." *Journal of Politics* 65: 1083–1096.

Green, Donald P., Mary C. McGrath, and Peter M. Aronow. 2013. "Field Experiments and the Study of Voter Turnout." *Journal of Elections, Public Opinion and Parties* 23: 27–48.

Greenberg, David H. and Burt S. Barnow 2014. "Flaws in Evaluations of Social Programs: Illustrations From Randomized Controlled Trials." *Evaluation Review* 38: 359–387.

Greenberg, David H., Donna Linksz, and Marvin Mandell. 2003. *Social Experimentation and Public Policymaking*. Washington, DC: Urban Institute Press.

Greenberg, David and Mark Schroder. 2004. *The Digest of Social Experiments*. Washington, DC: Urban Institute Press, 3rd edition.

Greevy, Robert, Bo Lu, Jeffrey H. Silber and Paul Rosenbaum. 2004. "Optimal Multivariate Matching Before Randomization." *Biostatistics* 5: 263–275.

Grépin, Kate A., William Jack, and James P. Habyarimana. 2015. "The Impact of Mobile Maternity and Transport Vouchers and SMS Test Messages: Experimental Evidence from Western Kenya". Presentation to NYU Wagner School, October 30.

Grose, Christian. 2010. "Priming Rationality: A Theory and Field Experiment of Participation in Legislatures." Presented at 3rd annual CESS-NYU Experimental Political Science Conference, February 5–6, New York. http://cess.nyu.edu/conferences/2-2010/papers/GroseNYU2010.pdf. Accessed September 1, 2015.

Grose, Christian. 2014. "Field Experimental Work on Political Institutions." *Annual Review of Political Science* 17: 355–370.

Guan, Mei and Donald P. Green. 2016. "Noncoercive Mobilization in State-Controlled Elections: An Experimental Study in Beijing." *Comparative Political Studies* 39: 1175–1193.

Gueron, Judith M. and Howard Rolston. 2013. *Fighting for Reliable Evidence*. New York: Russell Sage Foundation.

Habyarimana, James, Macartan Humphreys, Dan Posner, and Jeremy Weinstein. 2007. "Why Does Ethnic Diversity Undermine Public Goods Provision? An Experimental Approach." *American Political Science Review* 101: 709–725.

Hacking, Ian. 1988. "Telepathy: Origins of Randomization in Experimental Design." *Isis* 79: 427–451.

Hallsworth, Michael, John List, Robert Metcalfe, and Ivo Vlaev. 2014. "The Behavioralist As Tax Collector: Using Natural Field Experiments to Enhance Tax Compliance." Unpublished paper. Ideas: https://ideas.repec.org/p/nbr/nberwo/20007.html, Accessed August 1, 2015.

Halpern, David, Clive Bates, Geoff Mulgan, and Stephen Aldridge with Greg Beales and Adam Heathfield. 2004. *Personal Responsibility and Changing Behaviour: The State of Knowledge and its Implications for Public Policy*. London: Cabinet Office.

Halpern, David and Peter John. 2005. "Generating Social Capital with Schools and Citizenship Education: An Analysis of a Panel Survey of Young People." Unpublished paper.

Halpern, David, Owain Service and Behavioural Insights Team. 2015. *Inside the Nudge Unit: How Small Changes Can Make a Big Difference*. London: Penguin Random House.

Harden, Jeffrey J. 2013. "Multidimensional Responsiveness: The Determinants of Legislators' Representational Priorities." *Legislative Studies Quarterly* 38: 155–184.

Hartmann, George W. 1936. "A Field Experiment on the Comparative Effectiveness of 'Emotional' and 'Rational' Political Leaflets in Determining Election Results." *The Journal of Abnormal and Social Psychology* 31: 99–114.

Haskins, Ron and Greg Margolis. 2015. *Show Me the Evidence: Obama's Fight for Rigour and Results in Social Policy*. Washington, DC: Brookings Institute Press.

Hatry, Harry P. 1994. "Collecting Data from Agency Records." In *Handbook of Practical Program Evaluation*, edited by Joseph S. Wholey, Harry P. Hatry, and Kathryn E. Newcomer. San Francisco, CA: Jossey-Bass.

Haynes, Laura, Owain Service, Ben Goldacre, and David Torgerson. 2012. *Test, Learn, Adapt: Developing Public Policy with Randomized Controlled Trials*. London: Cabinet Office.

Haynes, Laura, Donald P. Green, Rory Gallagher, Peter John, and David J. Torgerson. 2013. "Collection of Delinquent Fines: An Adaptive Randomized Trial to Assess the Effectiveness of Alternative Text Messages." *Journal of Policy Analysis and Management* 32: 718–730.

Heckman, James. 1992. "Randomization and Social Policy Evaluation." NBER Technical Working Paper. Cambridge, MA: National Bureau of Economic Research.

Henderson, Mark. 2012. *The Geek Manifesto: Why Science Matters*. London: Bantam.

Hendra, Richard, Charles Michalopoulos, and Dan Bloom. 2001. *Three-Year Impacts of Connecticut's Jobs First Welfare Reform Initiative*. New York: MDRC.

Hendra, Richard, James A. Riccio, Richard Dorsett, David H. Greenberg, Genevieve Knight, Joan Phillips, Philip K. Robins, Sandra Vegeris, and Johanna Walter, with Aaron Hill, Kathryn Ray, and Jared Smith. 2011. *Breaking the Low-pay, No-pay Cycle: Final Evidence from the UK Employment Retention and Advancement (ERA) Demonstration*. Department for Work and Pensions Research Report No. 765. London: DWP.

Henegan, Carl. 2010. "How Many Randomized Trials Are Published Each Year?", http://blogs.trusttheevidence.net/carl-heneghan/how-many-randomized-trials-are-published-each-year. Accessed August 20, 2015.

Hess, Douglas R., Michael J. Hanmer, and David W. Nickerson. 2015. "Encouraging Local Bureaucratic Compliance with Federal Civil Rights Laws: Field Experiments with Agencies Implementing the National Voter Registration Act." Unpublished paper. http://www.douglasrhess.com/uploads/4/3/7/8/43789009/hess_hanmer_nickerson_may_2015_nvra_compliance.pdf. Accessed August 8, 2015.

H. M. Treasury. 2011. *The Magenta Book: Guidance for Evaluation*. London: HMSO.

Humphreys, Macartan. 2015a. "Reflections on the Ethics of Social Experimentation." Working Paper. World Institute for Development Economics Research. www.wider.unu.edu/publications/working-papers/2015/en_GB/wp2015-018/. Accessed August 5, 2015.

Humphreys, Macartan. 2015b. "More on Worms." http://www.columbia.edu/~mh2245/w/Worms_Again.pdf. Accessed August 6, 2015.

Humphreys, Macartan and Jeremy M. Weinstein. 2009. "Field Experiments and the Political Economy of Development." *Annual Review of Political Science* 12: 367–378.

Humphreys, Macartan, Raul Sanchez de la Sierra, and Peter van der Windt. 2012. *Social and Economic Impacts of Tuungane Final Report on the Effects of a Community Driven Reconstruction Program in Eastern Democratic Republic of Congo*. New York: Columbia University. http://www.oecd.org/countries/democraticrepublicofthecongo/drc.pdf. Accessed August 3, 2015.

Hutchinson, Dougal and Ben Styles. 2010. *A Guide to Running Randomized Controlled Trials for Education Researchers*. Slough: National Foundation for Educational Research.

Hyde, Susan D. 2010. "Experimenting in Democracy Promotion: International Observers and the 2004 Presidential Elections in Indonesia." *Perspectives in Politics* 8: 511–527.

Imai, Kosuke. 2005. "Do Get-Out-the-Vote Calls Reduce Turnout? The Importance of Statistical Methods for Field Experiments." *American Political Science Review* 99: 283–300.

Imai, Kosuke, Gary King, and Clayton Nall. 2009. "The Essential Role of Pair Matching in Cluster-randomized Experiments, with Application to the Mexican Universal Health Insurance Evaluation." *Statistical Science* 24: 29–53.

Isaakidis, Petros and John P. A. Ioannidis. 2003. "Evaluation of Cluster Randomized Controlled Trials in Sub-Saharan Africa." *American Journal of Epidemiology* 158: 921–926.

Issenberg, Sasha. 2013. *The Victory Lab: The Secret Science of Winning Campaigns*. New York: Broadway Books.

Iyengar, Shanto and Donald Kinder. 1987. *News That Matters*. Chicago, IL: University of Chicago Press.

Jacobsen, Christian Bøtcher and Lotte Bøgh Andersen. 2015. "Is Leadership in the Eye of the Beholder? A Study of Intended and Perceived Leadership Practices and Organizational Performance." *Public Administration Review* 75: 829–841.

Jakobsen, Morten. 2013. "Can Government Initiatives Increase Citizen Coproduction? Results of a Randomized Field Experiment." *Journal of Public Administration Research and Theory* 23: 27–54.

Jakobsen, Morten and Simon Calmar Andersen. 2013. "Intensifying Social Exchange Relationships in Public Organizations: Evidence from a Randomized Field Experiment." *Journal of Policy Analysis and Management* 32: 60–82.

James, Oliver and Alice Moseley. 2014. "Does Performance Information About Public Services Affect Citizens' Perceptions, Satisfaction, and Voice Behaviour? Field Experiments with Absolute and Relative Performance Information." *Public Administration* 92: 493–511.

James, Oliver, Sebastian Jilke, and Gregg Van Ryzin (eds.). Forthcoming. *Experiments in Public Management Research*. Cambridge: Cambridge University Press.

Jilke, Sebastian. 2015. *Essays on the Microfoundation of Competition and Choice in Public Service Delivery*. PhD Dissertation. Rotterdam: Erasmus University.

John, Peter. 2013. "Experimentation, Behaviour Change and Public Policy." *Political Quarterly* 84: 238–246.

John, Peter and Toby Blume. 2015. *Using Nudges to Increase Online Blue Badge Renewal – Testing the Effects Through a Randomised Controlled Trial for Essex County Council*. Report for Essex County Council.

John, Peter and Tessa Brannan. 2008. "How Different are Telephoning and Canvassing? Results from a Get Out the Vote Field Experiment in the British 2005 General Election." *British Journal of Political Science* 38: 565–574.

John, Peter, Elizabeth Evans, and Michael Sanders. 2015a. "Targeting Voter Registration with Incentives: A Randomized Controlled Trial of a Lottery in a London Borough." *Electoral Studies* 40: 170–175.

John, Peter, Oliver James, Alice Moseley, Liz Richardson, Matt Ryan, and Gerry Stoker. 2015b. "Testing Models of Citizenship to Promote Student Volunteering: A Large-Scale Field Trial of Endorsement Strategies." Output from the project, *Citizen Contribution to Local Public Services: Field Experiments in Institutions Incorporating Social Information ES/J012424/1*.

John, Peter, Sarah Cotterill, Alice Moseley, Liz Richardson, Graham Smith, Gerry Stoker, and Corinne Wales. 2011. *Nudge, Nudge, Think, Think: Experimenting with Ways to Change Civic Behaviour*. London: Bloomsbury Academic.

John, Peter, Elizabeth Evans, and Michael Sanders. 2014. "Targeting Voter Registration with Incentives: A Randomized Controlled Trial of a Lottery in a London Borough." Unpublished paper.

Jones, Rhys, Jessica Pykett, and Mark Whitehead. 2013. *Changing Behaviours: On the Rise of the Psychological State*. Cheltenham: Edward Elgar.

Kahan, Brennan C., Vipul Jairath, Caroline J. Doré, and Tim P. Morris. 2014. "The Risks and Rewards of Covariate Adjustment in Randomized Trials: An Assessment of 12 Outcomes from 8 Studies." *Trials* 15: 139.

Kahneman, Daniel and Amos Tversky. 1974. "Judgment under Uncertainty: Heuristics and Biases." *Science, New Series* 185: 1124–1131.

Kahneman, Daniel and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica* 47: 263–291.

Kahneman, Daniel, Paul Slovic, and Amos Tversky. 1982. *Judgment under Uncertainty: Heuristics and Biases*. New York: Cambridge University Press.

Karlan, Dean and Jonathan Zinman. 2010. "Expanding Credit Access: Using Randomized Supply Decisions to Estimate the Impacts." *Review of Financial Studies* 23: 433–464.

Kittel, Bernard, Wolfgang J. Luhan, and Rebecca C. Morton. 2012. *Experimental Political Science Principles and Practices*. Basingstoke: Macmillan.

Konnikova, Maria. 2015. "How a Gay Marriage Study Went Wrong." *The New Yorker*, May 22, http://www.newyorker.com/science/maria-konnikova/how-a-gay-marriage-study-went-wrong. Accessed September 1, 2015.

Kurtz, Karl. 2011. "Legislators as Lab Rats?", *The Thicket at State Legislatures* .http://ncsl.typepad.com/the_thicket/2011/05/legislators-as-lab-rats.html. Accessed September 1, 2015.

LaCour, Michael J. and Donald P. Green. 2014. "When Contact Changes Minds: An Experiment on Transmission of Support for Gay Equality." *Science* 346(6215): 1366–1369. Editorial retraction June 5, 2015: 1100.

Levitt, Stephen D. and John A. List. 2009. "Field Experiments in Economics: The Past, the Present, and the Future." *European Economic Review* 53: 1–18.

Lichtenstein, Sarah and Paul Slovic. 1971. "Reversals of Preference between Bids and Choices in Gambling Decisions." *Journal of Experimental Psychology* 89: 46–55.

Lin, Winston and Donald P. Green. 2016. "Standard Operating Procedures: A Safety Net for Pre-Analysis Plans." *PS: Political Science & Politics* 49: 495–500.

List, John. 2011. "Why Economists Should Conduct Field Experiments and 14 Tips for Pulling One Off." *Journal of Economic Perspectives* 25: 3–16.

Loewen, Peter, Daniel Rubenson, and Leonard Wantchekon. 2010. "Help Me Help You: Conducting Field Experiments with Political Elites." *The Annals of the American Academy of Political and Social Science* 628: 165–175.

Loewen, Peter, John Lior Sheffer, Stuart Soroka, Tamir Sheafer, and Stefaan Walgrave. 2014. "Are Politicians Better Decision Makers?" Unpublished paper. Available at https://sites.duke.edu/2014bmp/files/2014/10/Loewenetal.pdf. Accessed October 23, 2016.

Loewenstein, George. 1999. "Experimental Economics from the Vantage-point of Behavioural Economics." *Economic Journal* 109: 25–34.

Lund, Frederick Hansen. 1925. "The Psychology of Belief." *The Journal of Abnormal and Social Psychology* 20: 63–81.

McCall, William Anderson. 1923. *How to Experiment in Education*. New York: Macmillan Company.

McClatchey, M. W., S. J. Cohen, and F. M. Reed. 1992. "The Usefulness of Matched Pair Randomization for Medical Practice-based Research." *Family Practice Research Journal* 12: 235–243.

McClendon, Gwyneth. 2012. "Ethics of Using Public Officials as Field Experiment Subjects." *Newsletter of the APSA Experimental Section* 3: 13–20.

McClendon, Gwyneth. 2014. "Social Esteem and Participation in Contentious Politics: A Field Experiment at an LGBT Pride Rally." *American Journal of Political Science* 58: 279–290.

McClendon, Gwyneth. 2016. "Race Responsiveness, and Electoral Strategy: A Field Experiment with South African Politicians." *Journal of Experimental Political Science* 3: 60–74.

McCollum, Elmer V. 1924. "The Nutritional Value of Milk." *Proceedings World Dairy Congress* 1: 421–437.

McCord, Joan. 1978. "A Thirty-year Follow-up of Treatment Effects." *American Psychologist* 2: 284–289. http://www.childtrends.org/?programs=cambridge-somerville-youth-study#sthash.ZGAIgnhK.dpuf. Accessed September 1, 2015.

McDougall, Cynthia, Amanda Perry, Jane Clarbour, Roger Bowles, and Gillian Worthy. 2009. *Evaluation of HM Prison Service Enhanced Thinking Skills Programme*. London: Ministry of Justice Research Series 3/09. http://www.justice.gov.uk/publications/. Accessed September 1, 2015.

Mahéo, Valerie-Anne. 2014. "Can Voting Aid Applications Mobilize Citizens? A Field Experiment in a Low Income Neighborhood, in the Context of the 2014 Quebec Election." Paper given to the Workshop "Vision in Methodology", Hamilton, Canada, May 20–22.

Malesky, Edmund J., Paul J. Schuler, and Anh Tran. 2012. "The Adverse Effects of Sunshine: A Field Experiment on Legislative Transparency in an Authoritarian Assembly." *American Political Science Review* 106: 762–786.

Margetts, Helen Z. 2011. "Experiments for Public Management Research." *Public Management Review* 13: 189–208.

Maxfield, Myles, Laura Castner, Vida Maralani, and Mary Vencill. 2003. *The Quantum Opportunity Program Demonstration: Implementation Findings*. Princeton, NJ: Mathematica Policy Research.

Medical Research Council. 1948. "Streptomycin Treatment of Pulmonary Tuberculosis." *British Medical Journal* 2(4582): 769–782.

Mendez, Matthew S. and Christian R. Grose. 2014. "Revealing Discriminatory Intent: Legislator Preferences, Voter Identification, and Responsiveness Bias." *USC CLASS Research Paper* No. 14–17. Available at SSRN: http://ssrn.com/abstract=2422596. Accessed September 1, 2015.

Michaelides, Marios, Eileen Poe-Yamagata, and Jacob Benus Dharmendra Tirumalasetti. 2012. *Impact of the Reemployment and Eligibility Assessment (REA) Initiative in Nevada*. Columbia, MD: Impaq International.

Michalopoulos, Charles. 2005. "Precedents and Prospects for Randomized Experiments." In *Learning More from Social Experiments: Evolving Analytic Approaches*, edited by Howard S. Bloom, 1–36. New York: Russell Sage.

Miguel, Edward and Michael Kremer. 2004. "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities." *Econometrica* 72: 159–217.

Miller, Roy E., David A. Bositis, and Denise L. Baer. 1981. "Stimulating Voter Turnout in a Primary: Field Experiment With a Precinct Committeeman." *International Political Science Review* 2: 445–459.

Minton, Judith Haber. 1975. "The Impact of 'Sesame Street' on Readiness." *Sociology of Education* 48: 141–151.

Moher, David, Sally Hopewell, Kenneth F. Schulz, Victor Montori, Peter C. Gøtzsche, P. J. Devereaux, Diana Elbourne, Matthias Egger, and Douglas G. Altman. 2010. "CONSORT 2010 Explanation and Elaboration: Updated Guidelines for Reporting Parallel Group Randomized Trials." *British Medical Journal* 340: c869.

Moore, Graham F., Simon Murphy, Katherine Chaplin, Ronan A. Lyons, Mark Atkinson, and Laurence Moore. 2014. "Impacts of the Primary School Free Breakfast Initiative on Socio-economic Inequalities in Breakfast Consumption among 9–11-year-old Schoolchildren in Wales." *Public Health Nutrition* 17: 1280–1289.

Moore, Ryan T., and Keith Schnakenberg. 2015. *blockTools: Block, Assign, and Diagnose Potential Interference in Randomized Experiments Version 0.6–2*. https://cran.r-project.org/web/packages/blockTools/blockTools.pdf. Accessed September 1, 2015.

Morton, Rebecca B. and Kenneth C. Williams. 2010. *Experimental Political Science and the Study of Causality*. Cambridge: Cambridge University Press.

Moseley, Alice, Oliver James, Peter John, Liz Richardson, Matt Ryan, and Gerry Stoker. 2015. "The Effect of Social Information on Volunteering for Non-profit Organizations." Paper prepared for the 73rd annual MPSA Conference, April 16–19, 2015, held at the Palmer House Hilton in Chicago, IL, panel. Experiments on Participation and Civic Engagement.

Munnell, Alice, ed. 1987. *Lessons from the Income Maintenance Experiments*. Boston, MA: Federal Reserve Bank of Boston.

Neyman, Jerzy, with cooperation of K. Iwaskiewicz and St. Kolodziejczyk. 1935. "Statistical Problems in Agricultural Experimentation (with discussion)." *Supplement to the Journal of the Royal Statistical Society* 2: 107–180.

Nickerson, David W. 2006. "Forget Me not? The Importance of Timing in Voter Mobilization." Paper Presented at the Annual Meeting of the American Political Science Association, Marriott, Loews Philadelphia, and the Pennsylvannia Convention Center, Philadelphia, August 30–September 2.

Nickerson, David W. 2007. "Quality is Job One: Professional and Volunteer Voter Mobilization Calls." *American Journal of Political Science* 51: 269–282.

Nickerson, David W. 2008. "Is Voting Contagious? Evidence from Two Field Experiments." *American Political Science Review* 102: 49–57.

Nickerson, David W., Ryan D. Friedrichs, and David C. King. 2006. "Partisan Mobilization Campaigns in the Field: Results from a Statewide Turnout Experiment in Michigan." *Political Research Quarterly* 59(1): 85–97.

Niven, David. 2015. "Can Republican African Americans win African American Votes? A Field Experiment." Paper presented at the Annual Meeting of the American Political Science Association, San Francisco, September 3–6.

Nyhan, Brendon and Jason Reifler. 2014. "The Effect of Fact-Checking on Elites: A Field Experiment on U.S. State Legislators." *American Journal of Political Science* 59: 628–640.

Oliver, Adam, ed. 2013. *Behavioural Public Policy*. Cambridge: Cambridge University Press.

Olken, Benjamin A. 2007. "Monitoring Corruption: Evidence From a Field Experiment in Indonesia." *Journal of Political Economy* 115: 200–249.

Olken, Benjamin A. 2010, "Direct Democracy and Local Public Goods: Evidence from a Field Experiment in Indonesia." *American Political Science Review* 104: 243–267.

Olken, Benjamin A. 2015. "Promises and Perils of Pre-Analysis Plans." *Journal of Economic Perspectives* 29: 61–80.

Olken, Benjamin A., Junko Onishi, and Susan Wong. 2014. "Should Aid Reward Performance? Evidence from a Field Experiment on Health and Education in Indonesia." *American Economic Journal: Applied Economics* 6: 1–34.

Orr, Larry L. 1999. *Social Experiments: Evaluating Public Programs with Experimental Methods*. Thousand Oaks, CA: Sage.

Panagopoulos, Costas, Christopher W. Larimer, and Meghan Condon. 2014. "Social Pressure, Descriptive Norms, and Voter Mobilization." *Political Behavior* 36: 451.

Pasteur, Louis. 1881. "Summary Report of the Experiments Conducted at Pouilly-le-Fort, Near Melun, on the Anthrax Vaccination." *Yale Journal of Biology and Medicine* 75: 59–62.

Petrosino, Anthony, Carolyn Turpin-Petrosino, and John Buehler. 2003. "Scared Straight and Other Juvenile Awareness Programs for Preventing Juvenile Delinquency (Updated C2 Review)." In *The Campbell Collaboration Reviews of Intervention and Policy Evaluations (C2-RIPE)*, November. Philadelphia, PA: Campbell Collaboration.

Pocock, Stuart J. 1983. *Clinical Trials: A Practical Approach*. Chichester: John Wiley and Sons.

Pollock, Jon. 2006. "Two Controlled Trials of Supplementary Feeding of British School Children in the 1920s." *Journal Royal Society Medicine* 99: 323–327.

Pons, Vincent. 2014. "Does Door to Door Canvassing Affect Vote Shares: Evidence from a Countrywide Field Experiment in France." Working Paper: http://economics.mit.edu/files/9000. Accessed September 1, 2015.

Pons, Vincent and Guillaume Liegey. 2013. "Increasing the Electoral Participation of Immigrants: Experimental Evidence from France." Working Paper: http://economics.mit.edu/files/8885. Accessed September 1, 2015.

Putnam, Robert. 1993. *Making Democracy Work*. Princeton, NJ: Princeton University Press.

Riccio, James and Howard S. Bloom. 2002. "Extending the Reach of Randomized Social Experiments: New Directions in Evaluations of American Welfare-to-Work and Employment Initiatives." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 165: 13–30.

Richardson, Liz and Peter John. 2012. "Who Listens to the Grassroots? A Feld Experiment on Informational Lobbying in the UK." *British Journal of Politics and International Relations* 14(4): 595–612.

Robins, Philip K. and Robert G. Spiegelman, eds. 2001. *Reemployment Bonuses in the Unemployment Insurance System: Evidence from Three Field Experiments*. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

Rogers, Todd. 2014. "Low Cost/High ROI Experiments in Organizations." Unpublished paper.

Rogers, Todd and Joel Middleton. 2015. "Are Ballot Initiative Outcomes Influenced by the Campaigns of Independent Groups? A Precinct-Randomized Field Experiment Showing That They Are." *Political Behavior* 37(3): 567–593.

Ross, Heather L. 1970. *An Experimental Study of the Negative Income Tax*. Massachusetts Institute of Technology. Department of Economics. PhD Thesis.

Rothwell, Peter M. 2015. "External Validity of Randomized Controlled Trials: To Whom do the Results of this Trial Apply?" *The Lancet* 365: 82–93.

Rubin, Donald B. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66: 688–701.

Ryan, Matt, Gerry Stoker, Peter John, Alice Moseley, Oliver James, and Liz Richardson. 2015. "Remaking Heaven in Small-scale Democracy: A Randomized Experiment to Encourage Contested Elections and Greater Representativeness in English Local Councils." Unpublished paper.

Sanders, Michael and Aisling Ní Chonaire. 2015. "Powered to Detect Small Effect Sizes: You Keep Saying That. I Do Not Think It Means What You Think It Means." *Working Paper* No. 15/337, Centre for Market and Public Organisation. www.bris.ac.uk/media-library/sites/cmpo/documents/WP15337_Web_Version.pdf. Accessed August 1, 2015.

Sanders, Michael, Aisling Ní Chonaire, Veerle Snijders, and Peter John. 2015. "Non-Standard Matching in Charitable Giving – Null Results from Two Field Experiments." Unpublished paper. Available at http://38r8om2xjhhl25mw24492dir.wpengine.netdna-cdn.com/wp-content/uploads/2015/08/NSM-Paper-WP-1.pdf.

Schirm, Allen, Elizabeth Stuart, and Allison McKie. 2006. *The Quantum Opportunity Program Demonstration: Final Impacts*. Washington, DC: Mathematica Policy Research.

Schultz, P. Wesley, Jessica M. Nolan, Robert B. Cialdini, Noah J. Goldstein, and Vladas Griskevicius. 2007. "The Constructive, Destructive, and Reconstructive Power of Social Norms." *Psychological Science* 18: 429–434.

Schultz, Paul T. 2004. "School Subsidies for the Poor: Evaluating the Mexican Progresa Poverty Program." *Journal of Development Economics*. 74: 199–250.

Sekhon, Jasjeet S. 2010. "The Neyman-Rubin Model of Causal Inference and Estimation via Matching Methods." *The Oxford Handbook of Political Methodology*, edited by Janet Box-Steffensmeier, Henry Brady, and David Collier. Oxford: Oxford University Press.

Service, Owain, Michael Hallsworth, David Halpern, Felicity Algate, Rory Gallagher, Sam Nguyen, Simon Ruda, Michael Sanders with Marcos Pelenur, Alex-Gyani, Hugo Harper, Joanne Reinhard, and Elspeth Kirkman. 2014. *EAST: Four Simple Ways to Apply Behavioural Insights*. London: Behavioural Insights Team.

Shafir, Eldar, ed. 2013. *The Behavioral Foundations of Public Policy*. Princeton, NJ: Princeton University Press.

Shaw, Daron R., Donald P. Green, James G. Gimpel, and Alan S. Gerber. 2012. "Do Robotic Calls from Credible Sources Influence Voter Turnout or Vote Choice? Evidence From a Randomized Field Experiment." *Journal of Political Marketing* 11: 231–245.

Sherman, Lawrence W. and Dennis P. Rogan. 1995. "Deterrent Effects of Police Raids on Crack Houses: A Randomized, Controlled Experiment." *Justice Quarterly* 12: 755–781.

Sherman, Lawrence W. and David Weisburd. 1995. "General Deterrent Effects of Police Patrol in Crime Hot Spots: A Randomized, Controlled Trial." *Justice Quarterly* 12: 635–648.

Shu, Lisa L., Nina Mazar, Francesca Gino, Dan Ariely, and Max H. Bazerman. 2012. "Signing at the Beginning Makes Ethics Salient and Decreases Dishonest Self-reports in Comparison to Signing at the End." *Proceedings of the National Academy of Sciences* 109: 15197–15200.

Silva, Antonio and Peter John. 2014. *Improving Fees Collection at UCL: The Use of Behavioural Insights*. Unpublished report.

Sinclair, Betsy, Margaret M. McConnell, and Donald P. Green. 2012. "Detecting Spillover Effects: Design and Analysis of Multilevel Experiments." *American Journal of Political Science* 56: 1055–1069.

Skivington, Kathryn, Gerry McCartney, Hilary Thomson, and Lyndal Bond. 2010. "Challenges in Evaluating Welfare to Work Policy Interventions: Would an RCT Design Have Been the Answer to All Our Problems?" *BMC Public Health* 10: 254.

Smedslund, G., Kåre Birger Hagen, Asbjørn Steiro, Torill Johme, Therese Kristine Dalsbø, and Mons Georg Rud. 2006. *Work Programmes for Welfare Recipients*. www.campbellcollaboration.org/lib/download/88/. Accessed September 1, 2015.

Smith, Graham, Peter John, and Patrick Sturgis. 2013. "Taking Political Engagement Online: An Experimental Analysis of Asynchronous Discussion Forums." *Political Studies* 61: 709–730.

Soloman, Phyllis, Mary M. Cavanaugh, and Jeffrey Draine. 2009. *Randomized Controlled Trials: Design and Implementation for Community Based Psychosocial Interventions*. Oxford: Oxford University Press.

"Student" (W. G. Gossett). 1931. "The Lanarkshire Milk Experiment." *Biometrika* 23: 398.

Teele, Dawn Langan. 2015. "Reflections on the Ethics of Field Experiments." In *Field Experiments and Their Critics: The Uses and Abuses of Experimentation in the Social Sciences*, edited by Dawn Langan Teele, 115–140. New Haven, CT: Yale University Press.

Thaler, Richard H. and Shlomo Benartzi. 2004. "Save More Tomorrow: Using Behavioral Economics to Increase Employee Saving." *Journal of Political Economy* 112: 164–187.

Thaler, Richard H. and Cass R. Sustein. 2008. *Nudge: Improving Decisions about Health, Wealth and Happiness*. New Haven, CT: Yale University Press

Thomas, Duncan P. 1997. "Sailors, Scurvy and Science." *Journal of the Royal Society of Medicine* 90: 50–54.

Torgerson, David J. and Carole J. Torgerson. 2008. *Designing Randomized Trials in Health, Education and the Social Sciences*. Basingstoke: Palgrave.

Toth, Benjamin. 1989. "Clinical Trials in British Medicine 1858–1948, With Special Reference to the Development of the Randomized Controlled Trial." A dissertation submitted to the University of Bristol in accordance with the requirements of the degree of Doctor of Philosophy in the Faculty of Medicine. University of Bristol Department of Social Medicine, November 1998.

U.S. Department of Health, Education, and Welfare. 1976. *Summary Report: Rural Income Maintenance*. Washington, DC: Department of Health, Education, and Welfare.

Van de Vyver, Julie and Peter John. 2015. "Applying Behavioural Insights to Community Rights: A Randomized Evaluation of the Use of Norms to Achieve Behaviour Change in English Parishes." Unpublished paper.

Vicente, Pedro and Paul Collier. 2014. "Votes and Violence: Evidence from a Field Experiment in Nigeria." *Economic Journal* 124: 327–355.

Vivalt, Eva. 2015. "Heterogeneous Treatment Effects in Impact Evaluation." *American Economic Review: Papers & Proceedings* 105: 467–470.

Volpp, Kevin G., Leslie K. John, Andrea B. Troxel, Laurie Norton, Jennifer Fassbender, and George Loewenstein. 2008a. "Financial Incentive-based Approaches for Weight Loss: A Randomized Trial." *Journal of the American Medical Association* 300: 2631–2637.

Volpp, Kevin G., George Loewenstein, Andrea B. Troxel, Jalpa Doshi, Maureen Price, Mitchell Laskin, and Stephen E. Kimmel. 2008b. "A Test of Financial Incentives to Improve Warfarin Adherence." *Biomedical Central: Health Services Research* 8: 272.

Walker, Robert, Lesley Hoggart, Gayle Hamilton, with Susan Blank. 2006. *Making Random Assignment Happen: Evidence from the UK Employment Retention and Advancement (ERA) Demonstration*. DWP Research Report No 330. London: Department for Work and Pensions.

Wantchekon, Leonard. 2003. "Clientelism and Voting Behavior: Evidence from a Field Experiment in Benin." *World Politics* 55: 399–422.

Webb, E. J., D. T. Campbell, R. D. Schwartz, and L. Sechrest. 1966, *Unobtrusive Measures: Nonreactive Research in the Social Sciences*. Chicago, IL: Rand McNally.

White, Ariel R., Noah L. Nathan, and Julie K. Faller. 2015. "What Do I Need to Vote? Bureaucratic Discretion and Discrimination by Local Election Officials." *American Political Science Review* 109: 129–142.

White, Michael and Jane Lakey. 1992. *The Restart Effect: Evaluation of a Labour Market Programme for Unemployed People*. London: Policy Studies Institute.

Willis, Derek. 2014. "Professors' Research Project Stirs Political Outrage in Montana." *New York Times*, October 28, 2014.

Worthy, Ben, Peter John, and Matia Vannoni. 2015. "Information Requests and Local Accountability: An Experimental Analysis of Local Parish Responses to an Informational Campaign." Unpublished paper.

Yamanda, Kyohei and Alan Gerber. 2009. "Field Experiment, Politics, and Culture: Testing Social Psychological Theories Regarding Social Norms Using a Field Experiment in Japan." Paper presented at the Annual Meeting of the Midwest Political Science Association 67th Annual National Conference, The Palmer House Hilton, Chicago, IL, April 2, 2009.

Yates, Frank. 1964. "Sir Ronald Fisher and the Design of Experiments." *Biometrics* 20: 307–321.