Field Experiments in Comparative Politics and Policy

Donald P. Green and Peter John

The ANNALS of the American Academy of Political and Social Science 2010 628: 6
DOI: 10.1177/0002716209351498

Field Experiments in Comparative Politics and Policy

By
DONALD P. GREEN
and
PETER JOHN

If by some quirk or device—such as a time machine, capsule, or tunnel—we were transported back to 1998, a dozen years before this volume was produced, we would be thinking about field experiments in political science in a very different way than we do now. Although some early experiments had been conducted in field settings, by the 1980s, students of political behavior had largely abandoned them in favor of the analysis of large-scale surveys. Not a single field experiment was published in any political science journal in the decade preceding 1998, when field experimentation began to make a comeback. The first voting experiments, examining the impact of mail, telephoning, and canvassing on turnout, had been conducted but not yet published (Gerber and Green 2000). The next wave of field experiments in political science was just about to spread from Yale University to elsewhere in the United States and around the globe.

This renaissance set in motion two important changes in the discipline. These changes have occurred gradually, but a time traveler would now find them striking. One change had to do with the priority given to the rigorous estimation of causal effects. Prior to 1998, political scientists valued the enterprise of estimating cause-and-effect relationships, but they placed little emphasis on the role of research design,

Donald P. Green is the A. Whitney Griswold Professor of Political Science at Yale University. Since 1996, he has served as director of Yale's Institution for Social and Policy Studies, an interdisciplinary research center that emphasizes field experimentation. In collaboration with his Yale colleagues, he has conducted an array of field experiments in political science, communication, criminology, and education.

Peter John is the Hallsworth Chair of Governance and is a director of the Institute for Political and Economic Governance in the School of Social Sciences at the University of Manchester, UK. He is author of Analysing Public Policy *(Cassell 1998) and* Local Governance in Western Europe *(SAGE 2001).*

DOI: 10.1177/0002716209351498

especially experimental research design. Instead, researchers had grown accustomed to making strong and rather ad hoc assumptions in the course of proposing statistical models for causal inference. Another change had to do with the techniques by which experiments were carried out and analyzed. Having lost touch with experimentation and primary data collection in field settings, researchers who engaged in field experimentation had to acquire a new set of methods and skills. Anyone who has carried out a field experiment will attest to how complex the design and delivery of an intervention can be.

A mere dozen years later, the field has changed quite remarkably. Voter turnout experiments have expanded to investigate a wide range of mobilization campaigns, electoral contexts, and subgroups. The second edition of Green and Gerber (2008) identifies more than one hundred randomized trials and parallel experiments that have been conducted in other countries (e.g., John and Brannan 2008). The domain of subjects studied through field experimentation has expanded well beyond voter participation. Examples include the effect of media interventions on political opinions (Gerber, Karlan, and Bergan 2009), the salience of the content of candidate's campaign messages (Wantchekon 2003), the role of ethnicity in collective action (Habyarimana et al. 2007), and the influence of lobbying on legislative voting (Bergan 2009). As the subject matter has expanded, so too has the range of experimental settings. One now finds a vigorous field experimental agenda in a variety of social, economic, and cultural contexts, from North America and Europe to Africa, Asia, and Latin America.

In sum, there has been a profusion of experiments in real-world settings, and few in 1998 would have foreseen the diversity of interventions and experimental contexts that have come to characterize field experimental political science. The causal claims of the pre-1998 era have not necessarily been overturned, but they are now regarded in a different light. The advent of experiments with a high degree of internal and external validity has set a new standard of proof.

The essays commissioned for this volume reflect the growing maturity of field experiments in political science as researchers come to use the method in novel ways and tackle more subtle challenges of design and analysis. One set of challenges involves statistical inference. Although experiments are at heart simple methodological exercises, the real world presents a host of complications. Researchers do not always enjoy the cooperation of their subjects or of various agencies in charge of providing outcome measures. A burgeoning literature has developed to analyze the statistical consequences of untoward encounters between pristine experimental designs and the real world.

At an even more basic level, field experimenters often struggle to collaborate with public officials, nongovernmental organizations, and policymakers. This interaction works in a variety of ways. Experimenters sometimes rely on those who implement programs to randomize the treatment. Here, the intervention moves away from what the researcher can influence, such as a political campaign, to the more powerful and wider-ranging impact of the state in its various forms, such as giving out program grants or the varying ways it deals with citizens. This reliance creates the need for partnership with policymakers, which has advantages but

possible dangers, too. Another aspect of this nexus is the use of evidence from experiments. Academics may conduct research, but it is an open question whether non-academics will heed the findings that experiments produce. Then there are new directions for experiments, either in terms of procedures or domains, to expand further the field of experimentation, which generate further reflections on the method itself by those inside and outside the academy.

The contributions in this volume reflect these evolving concerns of experimenters. They derive from a conference of the title of this volume, held at the University of Manchester in the United Kingdom in July 2008. The conference was funded by the American Academy of Political and Social Science, the Institution for Social and Policy Studies at Yale University, and the University of Manchester Hallsworth Fund. It featured lively discussion among a wide array of experimental researchers, each of whom presented ideas about the promise and limitations of field experimental inquiry.

In the spirit of experimentation, we have randomized the author order in this volume using the random seed generator in Microsoft Excel. As a result, the contents are not structured by theme, but we can point out the groupings here. The first methodological piece is by Fernando Garcia and Leonard Wantchekon, who argue that theory-guided research agendas are likely to generate more efficient gains in experimental knowledge than an undirected inductive approach. Donald Green, Shang Ha, and John Bullock address the common criticism that experiments fail to illuminate the black box of causation. After showing the problem of drawing causal inferences about causal mechanisms using observational data, they show that the challenges remain formidable even with experimental designs. The black box critique is both unfair and unrealistic: only after substantial accumulation of experimental evidence can one hope to make secure inferences about the role of mediating factors.

A second category of methodological concerns has to do with the design and reporting of experiments. Hannah Ainsworth, David Torgerson, and Arthur Kang'ombe discuss how participants in a trial may be affected by disappointment about the experimental group to which they were assigned. The authors discuss a number of strategies for minimizing the bias that may occur when subjects refuse to participate in or drop out of studies because they prefer to be in a specific arm of the experiment. The behavior of human subjects in experimental settings is central to Marc Hooghe, Dietlind Stolle, Valérie-Anne Mahéo, and Sara Vissers's discussion of the use of student participants in experiments. To improve external validity, they argue for expanding the subject pool despite the practical challenges of doing so. Reporting conventions are central to the integrity of experimental science, and Isabelle Boutron, Peter John, and David Torgerson discuss the CONSORT statement that is a requirement for the publication of health care-related trials. CONSORT standards have improved the quality of randomized trials and have facilitated meta-analysis. The authors argue that a version of the checklist could improve the quality of reporting in political science. Finally, to conclude the methodological group of articles, Ana De La O and Daniel Rubenson discuss the problem where the units of measurement and randomization

do not overlap. They discuss and evaluate statistical strategies aimed at addressing this common feature of field experiments.

Several essays in this volume address the ways in which field experimenters interact with policymakers, public officials, and funding agencies. First, the piece by Dominic Pearson, David Torgerson, Cynthia McDougall, and Roger Bowles reports on how public agencies randomized the delivery of services using a "stepped wedge" design, whereby interventions are introduced at different points in time. Their surprising result is that randomization produced better implementation, the opposite of what policymakers generally expect. In contrast, Sarah Cotterill and Liz Richardson discuss the complexities of dealing with local policymakers as natural partners in setting up experiments, pointing out the difficulties researchers are likely to encounter. Peter Loewen, Daniel Rubenson, and Leonard Wantchekon examine how to deal with political elites when organizing and running political campaigns, paying particular attention to the ethical issues. They offer practical lessons from four case studies. In a similar vein, Gerry Stoker considers the difficulties experimenters face when communicating the results of their studies to policymakers, who have too many demands on their time to learn much from research. He argues that researchers need to be much more sensitive to the needs and perspectives of policymakers. Devra Moehler examines the relationships between funders, agencies, and researchers in the democracy and governance field. She reviews the growing number of these studies and discusses the complex environment for the delivery and use of these experiments.

The final set of essays suggests new types of experiments or aspects of experimental practice. Elizabeth Levy Paluck argues that qualitative outcome measurement can provide valuable information that may be of special importance to researchers who work with small numbers of observations. Susan Hyde's essay considers the range of possible domains to which field experiments apply. She suggests that experiments can be applied to fields that people would not initially consider, in this case to topics in international relations. Finally, Jane Green revisits the argument about the value of quasi-experiments as an alternative to field experiments, suggesting that they continue to be valuable when randomized controlled trials cannot be implemented effectively or quickly.

These essays are tributes to a developing style of research and the many creative scholars who are continually redefining the role that experimental methodology can play in the study of politics. Experimenters are grappling with an ever-expanding set of substantive research questions. At the same time, they need to ensure high standards in the generation and reporting of empirical results, perhaps even higher standards than the rest of political science. They seek to test the more advanced theories and produce results with the maximum possible internal and external validity. These essays are provocative and insightful contributions to that endeavor. If our time capsule were to return a dozen years from now, our hope is that the new generation of experimenters would have built on the insights in these essays as they expand the substantive and methodological frontiers of field experimentation.

References

Bergan, Daniel E. 2009. Does grassroots lobbying work? *American Politics Research* 37 (2): 327-52.

Gerber, Alan, and Donald P. Green. 2000. The effects of canvassing, direct mail, and telephone contact on voter turnout: A field experiment. *American Political Science Review* 94:653-63.

Green, Donald, and Alan Gerber. 2008. *Get Out the Vote: How to increase voter turnout*. 2nd ed. Washington, DC: Brookings Institution.

Gerber, Alan, Dean Karlan, and Daniel Bergan. 2009. Does the media matter? A field experiment measuring the effect on voting behavior and political opinions. *American Economic Journal* 1 (2): 35-52.

Habyarimana, James P., Macartan Humphreys, Daniel N. Posner, and Jeremy M. Weinstein. 2007. Why does ethnic diversity undermine public goods provision? *American Political Science Review* 101:709-25.

John, Peter, and Tessa Brannan. 2008. How different are telephoning and canvassing? A Get Out the Vote field experiment in the UK 2005 General Election. *British Journal of Political Science* 38:565-74.

Wantchekon, Leonard. 2003. Clientelism and voting behavior: Evidence from a field experiment in Benin. *World Politics* 55 (3): 399-422.