

Field Experiments and Their Critics

A Discussion of Dawn Langan Teele's *Field Experiments and Their Critics: Essays on the Uses and Abuses of Experimentation in the Social Sciences*

Field Experiments and Their Critics: Essays on the Uses and Abuses of Experimentation in the Social Sciences.
Edited by Dawn Langan Teele. New Haven, CT: Yale University Press. 280p. \$23.00.

Experimental approaches to political science research have become increasingly prominent in the discipline. Experimental research is regularly featured in some of the discipline's top journals, and indeed in 2014 a new *Journal of Experimental Political Science* was created, published by Cambridge University Press. At the same time, there are disagreements among political scientists about the limits of experimental research, the ethical challenges associated with this research, and the general model of social scientific inquiry underlying much experimental research. *Field Experiments and Their Critics: Essays on the Uses and Abuses of Experimentation in the Social Sciences*, edited by Dawn Langan Teele (Yale University Press 2015), brings together many interesting perspectives on these issues. And so we have invited a number of political scientists to comment on the book, the issues it raises, and the more general question of "the uses and abuses of experimentation in the social sciences."

Henry E. Brady

doi:10.1017/S1537592716003133

In *Novum Organum* (1620), Francis Bacon offered a new method of "true induction, with the assistance of experiment" (p. 117) based upon a series of tables (of presence, of absence, and of comparison) which were early versions of John Stuart Mill's "Joint Method of Agreement and Differences." Through his method he hoped to reveal "forms, . . . nothing else than those laws and regulations of simple action which arrange and constitute any simple nature" (p. 148). Bacon's use of his tables to understand the different forms of heat make amusing reading today (e.g., "rays of the sun," "all shaggy substances, as wool, contain some heat," and "horse dung, and the like excrement from other animals, when fresh" [pp. 121–22]), but his call for experimentation and induction helped to set western science on a new track away from scholasticism mired in syllogisms and reverence for Aristotle.

Henry E. Brady (hbrady@berkeley.edu) is Dean of the Goldman School of Public Policy and Class of 1941 Monroe Deutsch Professor of Political Science and Public Policy at the University of California, Berkeley.

Have the modern proponents of field experiments considered in this book found the New Atlantis and the true method of induction? Like Bacon, they offer us a recipe for the advancement of science: random assignment of cases to a treatment and control group, manipulation of a putative cause in the treatment group, and withholding the treatment from the control group to see what happens in the "counterfactual" world without the treatment. The result, they argue, will be reliable identification of true causal effects in real-world settings, just as Bacon promised. Will it?

The all-star cast of *Field Experiments and Their Critics* provide quite different answers to four basic questions about field experiments.

Is randomization and experimentation the only reliable way to do inference? Two articles—one by Alan Gerber, Donald Green, and Edward Kaplan (GGK) and the other by Abhijit Banerjee and Esther Duflo (BD)—argue that only experiments can reliably control for confounding factors that will bedevil and ultimately discredit observational studies. GGK present an interesting but "cooked" mathematical result that begins by assuming that experimental studies are unbiased while observational studies always suffer from some unknown bias. Not surprisingly, they conclude that experimental studies are better than

observational studies. BD run through all the ways experimenters can deal with their critics through the simple expedient of fielding even more experiments. Experimentation certainly has its virtues.

Are randomized control trials (RCTs) the gold standard, but perhaps usefully supplemented by other forms of research? Andrew Gelman calls RCTs the “gold standard” and plumps for “doing more field experiments,” but he goes on to say that “simple comparisons and regressions are and should be here to stay” (p. 185). He adduces two reasons. First even observational studies can be useful because “description is an important part of social science, both in its own right and in providing foundations upon which to build formal models” (pp. 185–86). This point deserves emphasis: Watson and Crick’s famous paper on the structure of DNA was merely description, but it was extraordinary science. Similarly, much of social science uses statistical methods (from the sample survey to regression analysis) to provide rich and reliable descriptions of the world that, just like Watson and Crick’s work, both suggest and restrict what is causally possible. Second, Gelman argues that it is wasteful and foolish to restrict the analysis of experimental data to just those analyses (often a simple difference of means test) justified by the randomized design. Certainly, Gelman argues, more information can be extracted by doing more statistical analysis, even if it is suspect.

Are experiments about as equally flawed as observational studies? Dawn Teele shows that experiments are gravely limited because they sometimes require manipulations that are unethical or undoable so that there is no way to avoid observational studies. Susan Stokes, Angus Deaton, and Ian Shapiro believe that experiments are flawed because they only provide “local average treatment effects” which do not take into account the heterogeneity in the world which means that individual cases might react quite differently to the same treatment. As a result, Stokes and others in the volume argue that the radical skeptic who distrusts all observational studies should be equally (radically) skeptical of experiments because their results can be biased. An experiment can show that on average a drug does not work because (unbeknown to the researcher) it interacts with some individual characteristic and makes half its recipients better and half of them worse off. Yet if we could identify the relevant characteristic, we could use the drug to improve people’s health. There is every reason to believe that heterogeneity like this is rife in social science. Finally, as Deaton argues forcefully, experiments inevitably require isolation and simplification that does not exist in the real world so that projecting a result beyond an experiment requires observational studies that consider general equilibrium effects, real-world implementation problems, and behavioral adaptations by people to changes in the world. In short, experiments may be undoable, misleading, or unprojectable.

Are there ways other than experimentation to do useful research? Stokes, Deaton, Kosuke Imai-Gary King-Elizabeth Stuart (IKS), and Christopher Barrett-Michael Cater (BC) certainly think so. Stokes states the essential issue: Radical skeptics believe that there are endless numbers of confounders that can undo observational studies, but she argues that the range of plausible confounders and interactions is not infinite. BC argue that substantive knowledge can rule out many possible confounding variables, and we should “retreat from radical skepticism and let theory and careful observation. . . guide an understanding of the causal process and name the potential confounds that can cripple inference from both observational data and from RCTs” (p. 76). Deaton agrees, and IKS provide a useful smorgasbord of alternative designs, tallying their strengths and weaknesses.

Where would Bacon stand on this? Would he applaud field experimenters as the true prophets of reliable induction? I don’t think so. Bacon depicted scientists as either “empirics or dogmatical,” and “the former like ants only heap up and use their store, the latter like spiders spin out their own webs” (p. 76). In this entomological competition, Bacon favored the bee who “extracts matter from the flowers of the garden and the field” and “works and fashions by its own efforts” (p. 76). The bee does not just collect facts or results like the pure empiric, and she does not simply spin out web after web of theoretical speculation like the scholastics. The bee extracts material from the world and refashions it. “The true labor of philosophy resembles hers, for it neither relies entirely or principally on the powers of the mind, nor yet lays up in the memory the matter afforded by the experiments of natural history and mechanics in its raw state, but changes and works it in the understanding” (p. 76). Bacon believed that scientists must link empirical data and theory to develop useful inductions. Science requires models and theory to discipline the ant-like heaping up of experimental facts. It requires facts to discipline the spider-like spinning out of theories. By extension, reliable empirical research requires both experiments and observational studies that blend empirics with theory. I’m with Bacon, science requires busy bees. And by the way, this is a honey of a book.

Yanna Krupnikov

doi:10.1017/S1537592716003145

Decisions about methodological approaches are pivotal to empirical political science research. Often, however, discussions of the costs and benefits of a given method take place in only a few paragraphs of a journal article. These paragraphs are often designed to highlight the benefits of and dismiss the common critiques against whichever method the author has used in a particular piece of research. By necessity these discussions are often brief and rarely examine how a methodological approach fits within the broader ideas of scientific research.

The relative brevity of “methods” sections is what makes books like *Field Experiments and Their Critics* useful. This volume, edited by Dawn Langan Teele, offers an in-depth, multi-perspective examination of a methodological approach. Pivotal to this examination is the second part of the title—*And Their Critics*. More than a discussion of the tremendous potential that field experiments hold for research, the book navigates through the possible costs and benefits of a method to which Teele refers in the book’s preface as an “experimental juggernaut” (p. vii).

Beginning with a nuanced chapter by Alan Gerber, Donald Green, and Edward Kaplan, “The Illusion of Learning from Observational Research,” the book raises a number of questions, some of which (methodological implementation, ethics, etc.) will ring familiar to scholars who use and are interested in experimental research. Other questions consider what it means to conduct research and propose theories in political science more broadly. Rather than discussing each chapter in isolation, I want to consider the book as the sum of its parts.

In bringing together a variety of scholarly perspectives, this volume attempts to shift scholars away from what I would call a “heuristic” approach to research. In simplest terms, relying on heuristics means using a limited number of cues to glean broader conclusions. If we use method as a heuristic, it can become a blanket cue about the overall usefulness or quality of a work—regardless of the way the method fits within the broader context of a scholar’s data, theory, and research question. The discourse across the chapters of this volume deliberately adds texture and conditionality to the evaluations of method, challenging the application of method as a heuristic. Taken as a whole, the chapters point to the interpretation of a method within the context of the scholar’s goals. Understanding the role of method within a context, this volume suggests, means not only considering the relationship between method and data or between method and theory, but

between method and a scholar’s underlying goal in asking a question.

Also notable is the book’s reframing of a number of debates surrounding field experiments in particular—but most experimental research in general. A novel approach to the debates surrounding field experiments is particularly apparent in the volume’s final chapter, titled “Methods Are Like People: If you Focus Only on What They Can’t do, You Will Always Be Disappointed.” In this chapter, Ian Shapiro takes a broader view of methods, writing about what it means to ask research questions. The points raised in this chapter are the types of considerations that are unlikely to be brought up in the “methods” section of a journal article. Grappling with the issues Shapiro raises (what, for example, makes a scientific result an important contribution?) could be a useful exercise even for those who disagree with Shapiro’s ultimate conclusions.

Although there is a commendable diversity of perspectives in this volume, one can’t help but wonder if the set of perspectives could have been even more diverse. It may have been interesting, for example, to include the voice of a scholar who relies largely on laboratory or survey experiments in the social psychological tradition. This type of method may often receive the sharpest of the critiques levied against experiments in political science and it may have been interesting to hear a scholar from this tradition grapple with the issues raised in this volume.

Overall, however, the usefulness of the volume as a whole is that it asks scholars to consider questions beyond “how can I convince a reviewer that this is a valid method?” It may be tempting for readers to select and cite the individual chapters that are most useful to their overall goal of reviewer persuasion. The benefit of the volume, however, is in the collection of ideas—even if those ideas do little to bolster the validity of one’s own methodological perspective.

Yanna Krupnikov (yanna.krupnikov@stonybrook.edu) is Associate Professor of Political Science at Stony Brook University.

Jessica Robinson Preece

doi:10.1017/S1537592716003157

Alan Gerber, Donald Green, and Edward Kaplan begin their volume, *Field Experiments and Their Critics*, with a forceful argument in favor of field experiments. They derive the “Illusion of Observational Learning Theorem” and then discuss the implications for the efficient allocation of research resources. They conclude that it generally makes sense to allocate all of one’s research resources to experimentation. Only when the costs (financial or ethical) of experimentation are extremely high and good-quality observational data is readily available should researchers invest in observational research. They throw down the gauntlet: Either you’re doing experiments, or you’re probably wasting your time.

The remaining authors mostly ignore the barb in this argument. None of them deny that field experiments can give us an excellent combination of internal and external validity—if they are appropriate and well-done. Hence, the core debate of the book is about the circumstances under which field experiments are appropriate and the extent to which they can be done well.

Abhijit Banerjee and Esther Duflo (Chapter 4) thoughtfully defend field experiments from several common critiques and outline how field experiments (mostly program evaluations) have advanced knowledge in development economics. But others in the volume are less convinced. Susan Stokes (Chapter 2) and Christopher Barrett and Michael Carter (Chapter 3) question how well experiments can actually be done and make note of problems with the interpretation of their results. Angus Deaton (Chapter 6) echoes these concerns and asks whether focusing on experimental methods causes us to focus on the wrong questions. Dawn Teele (Chapter 5) raises hard questions about ethics and the appropriateness of field experimentation, and though I do not agree with every point she makes, this chapter should be required reading for all field experimentalists.

The final three chapters try to mediate the conflict. Andrew Gelman (Chapter 7) touches on a number of important points: Uncovering causation isn’t the only reason to look at data, and randomized controlled trials aren’t the only research designs that use experimentation. Kosuke Imai, Gary King, Elizabeth and Stuart (Chapter 8) outline the ways in which observational and experimental researchers misunderstand each other and causal inference, and they offer helpful, practical advice. Finally, Ian Shapiro (Chapter 9) closes on an ecumenical note. All methods have strengths and weaknesses. Methodological pluralism is and always will be our best approach because it keeps the focus on the questions we are asking.

Even as a strong advocate for field experiments, Shapiro’s essay resonated deeply with me. Good questions form the foundation of social science research. Methods are crucial, but they must be secondary to questions or else we risk wasting time in a drunkard’s search. In this way, the observational vs. experimental debate echoes the old qualitative vs. quantitative debate. What is the best method? It depends on the question. Even Gerber, Green, and Kaplan ultimately give that answer, albeit somewhat grudgingly.

My main critique of the volume is that it neglects the practical aspect of field experiments that helps place them in methodological context. Randomized controlled trials are exceptional at answering the following question: Did this exact intervention change outcomes? That part is straightforward. The hard part about experimentation is designing interventions in such a way that the answer to the question is interesting regardless of what it is. Yes, it is technically true that you can learn *something* from any careful, well-powered, experimental intervention. But obviously the goal is to learn something *useful*.

I don’t bring this up to legitimize the flippant argument that one can only do experiments on topics that don’t matter. That is clearly not true. I bring it up because field experimentalists need to be honest about the debt they owe to observation as they design their interventions.

Pursuing an experimental research program is like engaging in a very expensive game of “Twenty Questions.” You present an experiment with a clear and precise hypothesis—a best guess generally rooted in existing observational research. If all goes well, the experiment answers back with a clear and precise yes or no. In the next round of the game, you know a little bit more (usually about what doesn’t work!), but you still have to come up with a new intervention and hypothesis. During this costly iterative process, experimentalists may be skeptical of observational findings, but they would be foolish to simply dismiss or ignore them.

In short, the best field experiments are creative and carefully implemented, but they are also well-informed by theory and cultural context. Theory and cultural context mostly remain the domain of observational research. Thus, taking an inclusive approach to methods is practical and wise for those engaging in field experimentation.

Jessica Robinson Preece (jessica_preece@byu.edu) is Assistant Professor of Political Science at Brigham Young University.

Peregrine Schwartz-Shea

doi:10.1017/S1537592716003169

In recent years, experiments—in laboratories, but especially in the field—have come into their own in political science and developmental economics, accompanied, often, by the argument that these methods and the evidence they produce are superior to others or, in the increasingly tired metaphor, that they constitute a “gold standard” to which all researchers should aspire. The implication is that researchers using other methods are doing “second-best” research and that funders and policy-makers should not waste their money or attention on studies conducted using “nonexperimental” methods. If accepted, such arguments directly contradict the value of methodological pluralism embraced by many political scientists post-Perestroika. This volume, edited by Dawn Langan Teele, speaks to whether field experiments themselves meet the “gold standard” and whether dominance by one method is preferable to methodological diversity.

As Teele writes in the preface, the book grew out of a debate she organized at Yale in 2009 in which Donald Green defended field experiments against criticisms by political scientists Susan Stoker and Ian Shapiro and development economist Angus Deaton. Teele represents the debate as one about the comparative value of experimental and “observational” inquiry in producing the best evidence for identification of causes in the social world. (In this context, “observational inquiry” means application of statistical or econometric methods to quantitative indicators or data sets, and this is its meaning here unless otherwise noted. This usage comes out of the research practices of epidemiology and public health; the work of statistician Paul Rosenbaum [*Observational Studies*, 2002] is the cited source for it.) The virtue of the book is that it gathers in one volume the arguments of some of the strongest proponents of field experiments in political science and developmental economics, alongside respectful assessments of this trend as well as some pointed critiques of it. Its Achilles’ heel is that only three of the chapters are original contributions, making it difficult to capture in print the responsiveness of a debate when six chapters are reprints of already published work.

Why *field* experiments? The conventional wisdom is that laboratory experiments are strong on the criterion of internal validity (the identification of causal effects, i.e., whether and to what extent manipulation of the independent variable, the “treatment,” produces the dependent variable, the effect) but weak on external validity (generalizability of findings to settings outside the lab). In contrast, observational methods are weak on internal

validity (due to the inability to manipulate the independent variable and measure its average effects using equivalent experimental and control groups) but strong on external validity (since the observations are measures of the world). It follows, in this view, that field experiments are strong on both criteria, employing the power of random assignment to produce equivalent control and experimental groups for the identification of causes, yet more generalizable because “the field” is not artificial like “the lab.”

Field Experiments and Their Critics opens with an example from political science, a reprint of Alan Gerber, Donald Green, and Edward Kaplan’s 2004 “The Illusion of Learning from Observational Research,” in which they offer a Research Allocation Theorem for deciding when to conduct experimental versus observational research. As the title implies, they argue, based on their theorem, that experiments are usually superior but are, also, a standard in and of themselves: “The test of whether methodological inquiry succeeds is its ability to correctly anticipate experimental results because experiments produce unbiased estimates regardless of whether the confounders [i.e., potential control variables] are known or unknown” (p. 25). Methodologists of other stripes might well question this circular logic and point to understanding or prediction of substantive political events as a better test.

The second chapter, by Susan Stokes, is “A Defense of Observational Research.” Although the problem of “omitted variable bias” is well recognized in the statistical approaches used in observational research, Stokes argues that the position that *only* random assignment controls for both known and unknown factors is a position of “radical skepticism” that caricatures observational researchers as able to do “nothing more than ‘assume nonconfoundedness’” (p. 38). One difficulty she identifies in the argument of Gerber, Green, and Kaplan is that claims about methods in the abstract often look quite different when substantive issues are brought into play. In observational research, there is often a limited number of alternative explanations (of the outcome of interest), many of which can be eliminated using substantive knowledge and theory. Stokes appears to have persuaded economists Christopher Barrett and Michael Carter, who conclude in the next chapter, “A Retreat From Radical Skepticism,” that development economists should “more creatively balance observation and experimental data” (p. 72). Describing weaknesses of field experiments—from ethical concerns to the ways in which participants may actively resist random assignment to control or experimental groups—they write: “[L]imits to RCTs [randomized controlled trials] . . . by themselves mandate a return to methodological pluralism if we are [to] continue to answer the important questions” (p. 74).

Chapter 4 is a reprint of Abhijit Banerjee and Esther Duflo’s 2009 “Experiments in Development,” which

Peregrine Schwartz-Shea (psshea@poli-sci.utah.edu) is professor of political science at the University of Utah.

recounts the successes of field experiments in Africa, Mexico, and India. Teele includes development economics in the book because of the “experimental revolution” in that field (p. 5), which has garnered much academic and public attention, specifically through the work of this duo who wrote the 2011 prize-winning book *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*. Although their essay responds to some “concerns” about experiments—randomization bias and compliance issues, among others—they do not believe these problems are unique to field experiments (p. 78). They end the essay by opining that economists’ insights “should guide policy making” and even “midwife the process of policy discovery” (p. 113).

Teele’s own chapter, up next, examines the ethics of field experiments. Unlike other methods, field experiments are akin to “social engineering,” in her view, because they require not only observations of people’s daily lives but also purposeful interventions that can alter individuals’ “life chances” and even undermine the “social fabric” of communities (pp. 135, 115, 129). Examining the work of Banerjee and Duflo and others in developmental economics, Teele argues that too many field experiments have violated the principles of the *Belmont Report* (1978), the foundational document in medical ethics referenced by U.S. institutional review boards.

Chapter 6, another reprint, is Angus Deaton’s detailed critique of field experiments as a primary method for understanding economic development. He is skeptical that field experiments as actually implemented are superior to the observational methods of econometrics. Although parts of his analysis are specific to technical debates in development economics, what is most germane to the book’s larger themes is that Deaton sees a pernicious effect from the dominance of field experiments: a focus on the “what” to the neglect of the “why,” which undermines the explanatory value of development economics. Like Stokes, Deaton is skeptical of the general proposition that RCTs “automatically trump other evidence” (p. 143), averring that there is “no substitute for careful evaluation of the chain of evidence and reasoning by people who have the experience and expertise in the field” (pp. 179–80). Another vote for moderation comes in Chapter 7 from Andrew Gelman, who, although he repeats the gold standard metaphor, finds space for such methods as historical and qualitative approaches and experimental approaches besides RCTs.

While its title appears on point, “Misunderstandings Between Experimentalists and Observationalists About Causal Inference,” the penultimate chapter, a reprint of Kosuke Imai, Gary King, and Elizabeth Stuart’s 2008 article in the *Journal of the Royal Statistical Society* (vol 171: 2, pp. 481–502), is written for a technically proficient audience. As with the authors of the beginning chapter, these scholars put their energy into a “general framework

for understanding causal inference” (p. 196) illustrated through the comparison of two studies on the survival of women with breast cancer. Although the general framework may correct “misunderstandings,” it is not clear whether other substantive disagreements between the two communities remain.

The final chapter, by Ian Shapiro, echoes the views of Deaton in its criticism of the disciplinary effects of too much emphasis on field experiments. Shapiro criticizes the effort to mimic practices from medicine, emphasizes the kinds of questions of interest to political scientists that are not amenable to investigation via field experiments, and worries that research agendas focused on the latter method mean that “scholars will be learning more and more about less and less” (p. 233). In a way, the volume itself is testimony to his concerns as there is relatively little attention to substantive politics of, for example, breast cancer research. Banerjee and Duflo write—naively, in my view—that the nonparticipation that can undermine field experiment logic will become less common in developing countries as randomized evaluation of development programs comes “to be recommended by most donors” (p. 97), ignoring the politics of international aid that scholars of international relations (nowhere mentioned) point to as important for understanding development.

Shapiro is now well known for his statement that became the title for his chapter in this volume: “Methods Are Like People: If You Focus on What They Can’t Do, You Will Always Be Disappointed.” Turning that part of his title around and focusing on what methods *can* do, those endorsing methodological pluralism will recognize that field experiments *are* good for some research questions—identifying and providing precise evidence for particular “treatment effects” of interest to policymakers. But consistent with this position, other methods have their own strengths, and it would have been nice to see some recognition of that in this volume. Contemplating the entirety of *Field Experiments and Their Critics*, many readers will likely demure from Teele’s assessment that “the jury is still out as to whether experiments are the only way or the best way to tell us all we need to know about a policy intervention” (p. 116). Her volume provides the arguments to send that jury home.

Betsy Sinclair

doi:10.1017/S1537592716003170

This book reflects a magnificent debate between highly regarded, serious researchers. It provides space for the authors to develop careful arguments in each chapter and the chapters provide a coherent sense for the weaknesses of experimental work circa 2009. As a greater preponderance of the chapters are leery of experimental work, this book could be a critical read for those scholars who do want to engage an experimental toolbox. It could be incorporated with a series of other readings such as Paul Rosenbaum's *Observational Studies*, Alan Gerber and Donald Green's *Field Experiments*, and papers by pro-experimentalists such as Guido Imbens ("Better Late Than Nothing") to produce an interesting course on research design.

The arguments in the book are slightly dated, as the original conversation began from a debate that arose within the Yale University social science community in 2009. Seven years later, there is more consensus within the experimental community about the role experiments play in addressing a research question. For example, experimentalists have formed a professional organization titled Experiments in Governance and Policy (EGAP) which houses a study registry, allowing experimentalists to test interactions in a more principled way. Additionally, a growing number of researchers have studied ways to incorporate social networks into randomized experiments, addressing concerns voiced by Susan Stokes that "there can be no stopping rule, no point at which one has tested for all possible interactions" (p. 49). Indeed, the marriage of "radical skeptics" (Stokes' term for the advocates of an experimental agenda above an observational agenda) to traditional political science has generated a host of new tools to address many of the concerns espoused in these chapters. The quality of inferences from experimental work has been increasing, moving to develop new statistical techniques to address some of the threats to causal inference pointed out in various chapters (i.e., heterogeneity, interference). That is, I would argue, "radical skeptics" have significantly improved our ability to draw inferences. Many more scholars are aware of the causal inference techniques, for example, described in Chapter 8 "Misunderstandings Between Experimentalists and Observationalists about Causal Inference."

This is because the academic community is increasingly concerned about causal inference. In their chapter Alan Gerber, Donald Green and Edward Kaplan state that, "the risk of bias is typically much greater in observational research" (p. 11) and indeed while this is true, there has been a growing body of literature that

has leveraged natural randomization (lotteries, natural disasters, etc.) to seize the opportunity to produce a causally-identified parameter from what is otherwise an observational study. In the discussion of experimental work, Dawn Teele emphasizes "experiments" as a methodology and writes with some concern about whether experimental researchers are inappropriately evangelizing and colonizing the discipline (i.e. "experimentalists' tracts" p. 7). This mischaracterizes the adoption of experimental work: it's proselytizers testify to the importance of causal identification, with randomized trial experimental work seen as one component of a revolution that recognizes endogeneity as one of the critical challenges of human subjects research.

The book is successful at bringing the issue of ethics to the table; an issue that experimentalists are still wrestling with and have yet to form a consensus on. As Teele describes carefully in her chapter on the ethics of field experiments, the current framework for evaluating whether an experiment is ethical is insufficient. Institutional Review Boards (IRB) are focused only on protecting subjects, not non-subjects. Teele's illustration of this point engaged Marianne Bertrand et al.'s (2009) experiment on bribes for the procurement of illegal driver's licenses in India. In this instance, the IRB did not consider whether the experiment, by providing an incentive for unqualified drivers, might harm other drivers or pedestrians. Experimental social science surely needs a set of more clearly delineated principles for the ethics of experimental work. These include cases where the current restrictions should probably be significantly relaxed, such as those where academics are working in partnerships with organizations or companies who are already engaged in experimental work. They also include cases where there should be more careful consideration of the broader impacts of experimentation, such as those instances where voter mobilization or persuasion experiments could affect the outcome of an election. Teele offers one particularly interesting suggestion, "to bring the participants in on the study's purpose after the trial is over and to share research results after the report is finished" (p. 136). This could be accomplished by providing ungated access to papers published from randomized field experiments, for example.

The book highlights the central tension between authors: good research questions versus good research designs. The very best research in our field is able to find good questions to answer with research designs that provide some measure of causal identification. Yet, there are many instances where researchers must choose to emphasize one over the other, and in that case there are real stakes. Observational work, as traditionally conducted, has real limitations where there is very likely bias introduced from endogeneity. Biased findings can be totally wrong. Wrong findings can cause great harm and lead to poor social policies. The experimentalists provide

Betsy Sinclair (bsinclair@wustl.edu) is Associate Professor of Political Science at Washington University in St Louis.

a strategy to mitigate impact from “unobservable confounders” (p. 35), even when we allow ourselves to be “radical skeptics” and limit the scope of their power. The alternative, however, of choosing only causally identifiable questions is also limiting. We might sometimes prefer to study questions where we adopt heroic assumptions and risk bias rather than risk not answering these important questions. Medical science has sometimes made this choice; imagine a world where we had never studied the relationship between smoking and lung cancer, or children’s health and lead paint, because of the infeasibility of conducting a randomized trial. We have similar issues in the social sciences as well. As Christopher Barrett and Michael Carter write, “macroeconomic and political economy questions that many believe to be of first-order importance in development are clearly not candidates for randomization” (p. 73).

The punch line of this book might be: building positive knowledge is tough. Each of these authors takes

seriously the charge that establishing meaningful findings in research needs careful and well-considered research design. Indeed, it appears there is more consensus in these chapters than would initially appear, and that many of these authors would agree with Stokes when she writes that “we need rich and variegated evidence, rigorously developed and analyzed, and considered in light of theories—which, as in all fields of science, are in part deductive in nature—if we are to gain knowledge about the workings of the social world” (p. 54). While experimental work has gained a strong foothold in political science, there are still those who have deep misgivings about its role. This book brings many of these concerns to the table in a clear and coherent fashion. From 2009 to 2016, we have seen experimental work play an increasingly prominent role in political science. This book challenges us to continue to think critically, radically, and skeptically about the role of experiments and the need for causal identification in research design.