I

Why a Primer on Social Science Experiments?

In the 1909 American Political Science Association’s presidential address, A. Lawrence Lowell stated, “We are limited by the impossibility of experiment. Politics is an observational, not an experimental, science” (Lowell 1910: 7). One hundred and ten years later, the association’s president, Rogers Smith (2020: 15), raised the question of “whether an excessive emphasis on experiments will unduly constrict the questions political scientists ask.” Clearly, much has changed in political science.¹ The same can be said about many social science disciplines where experiments have evolved from a nonexistent method to an accepted method to a primary method. Even psychology – where experiments have forever been a central approach – has experienced substantial changes in the last decade due to shifts in the social sciences. Specifically, massive technological advances have facilitated data access and analysis, and social scientists from all disciplines have called for more “open science” practices that involve transparency and replication. There is a concern, however, that

¹ While not my focus, another contrast between Lowell and Smith concerns their perspectives on race. Lowell served as the president of Harvard from 1909 to 1933, during which time he attempted to ban Black students from living in freshman halls (Sollors et al. 1993). In contrast, a fair amount of Smith’s work explores the incorporation of minorities into political life, such as pointing out that the United States has an ascriptive tradition that involves sexism, racism, and nativism (Smith 1993). He states in his presidential address, “we are not going to be able to understand major political developments of the past, present, and future if we do not explore more deeply the politics of identity formation, using all methods that can help” (12). This comparison reveals the extent to which the discipline has changed from one that largely ignored race for the first part of its history (Blatt 2018) to one that is now recognizing the central role of studying race and identity.
these changes may cause experimentalists to become “methods driven,”
neither asking appropriate questions nor maximizing the potential of the
method (e.g., Telen and Mahoney 2015: 19). These apprehensions
accentuate the need for careful discussion of the experimental method.
That is the goal of this book: The hope is to provide readers with a way to
“think” about experiments, both as users and as consumers.

From my perspective as a political scientist, I discuss the evolution of
experiments in political science and use many examples from the disci-
pline. That said, the arguments I make and the suggestions I offer apply to
any social science application of an experiment – which as will be dis-
cussed in detail – I define as a study where an intervention (by a researcher
or a natural event) provides the primary mechanism by which one
attempts to make a casual claim. What follows can be read by those with
no background and/or interest in political science.

To be clear, this book is not a vigorous defense of experiments,
although it will become apparent that experiments have far-reaching
applications. Further, the book is neither a textbook on experimental
design and analyses – many such treatments exist – nor an advanced
discussion of new developments, which is available in Druckman and
Green’s (2021) edited volume. Instead, the book is advisory and caution-
ary. As social scientists forge ahead with experiments, it is crucial they do
so in the most productive and careful manner possible, remembering what
experiments are for; why and when they should be used; and how they
should be designed, implemented, and evaluated. In short, this book will
help social scientists think about experiments more productively.

My argument can be summarized as follows: In some sense, the social
sciences have become fast moving; computing technology and easily
accessible data sources have led to an explosion of experiments. This is
a dramatic change from earlier generations where the central challenges
for experimentalists involved gaining access to computing power and
data. Today, experiments seem often to be designed and implemented
quickly and not connected to the full scientific process. I argue that the
result is a problem. Specifically, experiments need to be thought of as one
part of a scientific process and not the first part. They need to be used
when appropriate and build on/have an interplay with questions, observa-
tions, and theory. Moreover, conducting a quality experiment requires
thinking through a litany of decisions, ranging from how to think about
problems of casual inference to considering various counterfactuals to
how to interpret replications. A good experiment is slow moving (given
the host of considerations), which is counter to the current fast-moving
temptations available in the social sciences. This book is about thinking through the parts that make an experiment slow moving. I make this argument by discussing the following:

- How to think about the place of experiments in the social scientific process – that is, addressing the question of what role experiments play in the accumulation of knowledge (e.g., relative to theory building and other steps in the scientific method).
- How to arrive at questions that experiments are useful to answer.
- How to think about the assumptions underlying different types of experiments.
- How to think about evaluating the realism and validity of experiments, as well as assessing experimental samples.
- How to think about new experimental designs.
- How to proceed after an experiment is completed, and particularly how to think about the replication of experiments.
- How to think about the process of designing and conducting a “good” experiment; by this I do not mean the technical design details but rather all of the steps one should take to ensure experiments connect to theory and advance knowledge.

My approach seeks to broaden conversations about experiments by placing them in the larger research process where one must consider issues that ostensibly have little direct connection to experimental design (e.g., questions of sampling and measurement) but are essential if one hopes to design optimal experiments. At the same time, it focuses discussion by highlighting the need to attend to precise causal inference assumptions and counterfactual thinking. Further, I offer somewhat contrarian views on experimental realism and validity – perspectives that also lead to some cautionary notes when it comes to open science practices such as preregistration and replication. Some highlights of the points I make include the following.

In so doing, I hope to make clear that, counter to Smith’s (2020) assessment, experiments when done carefully need not constrict their reach. However, I simultaneously emphasize Smith’s point that experiments have a place in the scientific process and the key to exploiting their power is to understand that place, including their limitations. Put another way, I follow Smith’s (2020: 16) advice to “find ways to place our particular studies more explicitly in broader accounts of politics that can credibly indicate their importance.” Smith (2020: 15) further states that the “contributions of this experimental turn are undeniable.” My hope is that this book will help make them even greater.

Why a Primer on Social Science Experiments?
Experiments are useful only if there exists a substantively grounded question, a well-defined target population, carefully constructed measures, and clear points of comparisons. Many extant experiments fail to explicitly consider these issues (Chapter 2).

All experiments – whether using random assignment or relying on experimental control – involve assumptions about causal inference that often receive scant consideration (Chapter 2).

The goal of an experiment is to generalize a causal relationship. In many, but not all cases, the size of the effect from a single experiment is less important, and the sample used to document the effect is not crucial (Chapter 3).

In most cases, assessing whether experimental treatments resemble the “real world” is misguided, as the focus should be on the theoretical construct of interest and ensuring successful delivery of the treatments (Chapter 3).

Generalizing an experimental result is more complicated than evaluating the “representativeness” of the sample. In fact, the representativeness of the sample only matters when causal effects differ across relevant people (or when the goal of the experiment is to directly inform policy) (Chapter 3).

Recent design innovations that use audit field experiments and conjoint survey experiments offer many opportunities, but these designs have limitations and are only useful under particular circumstances. Ultimately, an experimental design is only as good as the question being addressed and the hypotheses being tested (Chapter 4).

The process of asking good questions for experiments develops from assessing the world and the field, socializing with a diversity of people, and building on prior experiments that did not go as planned (Chapter 5).

Implementing a good experiment requires the documentation of every decision in detail. However, it does not require a formal registration of a pre-analysis plan. Such a plan, if done, should not constrain an experimentalist from exploratory data analyses or incorporating theoretical ideas that had not been initially considered (Chapter 5).

After an experiment is done, repeating it for replication’s sake has limited value; however, replication can be used as a route to innovation (by extending prior designs) and aggregation so as to isolate the size of an effect (Chapter 5).

Despite all the innovations in experimental social science, the steps needed to design a quality experiment remain the same and require
Why a Primer on Social Science Experiments?

situating experiments in the entire scientific process. This starts with asking a relevant substantive question and, from there, involves a lengthy iterative process, but one that is doable and rewarding (Chapter 6).

Who would want to read this book? I hope the material is relevant to any social scientist, including students who are just learning about social science. Those who regularly conduct experiments may find that some parts are familiar and other parts provide novel views. Those who do not engage in experimentation, or even begrudge experiments, may learn about the logic of experimentation, novel applications, and/or how to interpret and generalize experiments. These have become requisite skills for reading social science literature.

I proceed in this chapter with a discussion of the evolution of experiments, illustrating this development through the field of political science. I argue that the discipline currently finds itself in a new era, parts of which apply to all of the social sciences. This new era began around 2010 and reflects the confluence of experiments achieving widespread acceptance in the discipline, technological advances, and the open science movement (these latter two dynamics have affected all of the social sciences). The era introduced many opportunities but also new challenges. Ironically, the ease of conducting experiments today has the potential to undermine their quality.

THE EVOLUTION OF EXPERIMENTS

In their foundational text on quasi-experiments, Campbell and Stanley (1963: 3) explain that we must justify experimentation on more pessimistic grounds – not as a panacea, but rather as the only available route to cumulative progress. We must instill in our students the expectation of tedium and disappointment and the duty of thorough persistence … We must expand our students’ vow of poverty to include not only the willingness to accept poverty of finances, but also a poverty of experimental results.

This pessimistic portrayal reflects the prevailing reality of experiments for much of social science history: Experimentalists had to overcome the logistical challenges of and limited opportunities for data collection. For example, in the first random assignment experiment published in the American Political Science Review (APSR), Eldersveld (1956) relied on fifty students and four staff members to work for about 400 hours.
to study 500 subjects – that is, it was far from a straightforward process. Iyengar et al.’s (1982) seminal agenda-setting experiment had all of twenty-eight subjects, while Druckman’s (2001) study of framing had 264 subjects but took roughly five months to collect the data. In addition to the shortage of readily available experimental subjects (even student subject pools can be used for only so many experiments), experimentalists also faced the inevitable occurrence of null results. These results were rarely published, which led Ioannidis (2005) to famously claim that “most published research findings are false.”

In the last few decades, experimentation has dramatically changed. Data collection opportunities are plentiful thanks to crowdsourcing platforms, internet panels, social media contacts, and elite samples via e-mail. Computing advances allow for large-scale experiments, sometimes on literally millions of participants (e.g., Bond et al. 2012). Moreover, scholars no longer dismiss null results as inherently uninteresting due to the recognition that publishing only statistically significant results can skew the research record. These developments bring with them new opportunities but also a new type of possible poverty. The ease of data collection and acceptance of non-findings means scholars might be less incentivized to design and implement quality experiments: There is much less at stake with each experiment, given the relative ease of data collection and increasing acceptance of null results. On the latter, it has become essential to distinguish meaningful null results out of a carefully constructed and implemented experiment from those resulting from a poorly designed study. In short, the concerns are a poverty of poor designs, inappropriate analyses, limited use of data, and/or flawed interpretations. Even an infinite amount of data cannot compensate for a thoughtlessly designed experiment. This makes it all the more important to ensure that experimentalists design sound studies and properly analyze, interpret, and present the data from particular samples. To situate the relevance of the aforementioned concerns, I next turn to an overview of how political science (as an example) arrived at its current state when it comes to experiments.

The Expansion of Experiments in Political Science

As mentioned, the lessons that follow apply to any social science discipline, but in this section I offer an example of how experiments have

3 Parts of this section and the next one come from Druckman and Green (2021).
emerged in my home discipline of political science. Similar trends, albeit at different points, have occurred in other disciplines such as economics and sociology. This differs from psychology where experiments have forever been central. However, as I will discuss, technological changes and the emergence of open science have altered experiments in all disciplines including psychology.

When it comes to political science, the Lowell quote with which I started the book makes clear that experiments were not present when the discipline launched. With a few notable exceptions – such as Gosnell’s (1926) study of voting mobilization – experiments remained, at best, peripheral through most of the twentieth century. The 1950s and 1960s saw some activity with a research program that used role-playing experimental simulations to test how situational factors affect decisions to go to war and international negotiations (e.g., Hermann and Hermann 1967; Mahoney and Druckman 1975; Guetzkow and Valadez 1981). A bit later, a short-lived journal titled *The Experimental Study of Politics* appeared. The status of experiments began to notably change, however, in the late 1980s and 1990s with experiments on Congressional committee decision-making (Fiorina and Plott 1978), media effects (Iyengar and Kinder 1987), elections (McKelvey and Ordeshook 1990), and public opinion (Sniderman et al. 1991) (also see Kinder and Palfrey 1993).

This slow emergence can be seen by charting the number of experimental articles in the discipline’s flagship journal, the *American Political Science Review* (APSR), as documented in Figure 1.1. There were no experimental articles from 1906 to 1956, and thus the figure begins in the 1950s and continues through 2019, reporting the number of experimental articles by decade.4 (This is not a cumulative count of articles but rather the specific number by decade.) The figure reveals the aforementioned bump in the 1980s and 1990s and also shows the continuing increase in subsequent decades to 31 in 2000–2009 and 84 in 2010–2019.5 The figure supports

---

4 For the content analysis, I use a broad definition of “experiment” that includes experiments that rely on random assignment; that pay subjects based on their behaviors; and that exploit random or near-random natural variations (i.e., “natural experiments”). I offer a formal definition of experiment in Chapter 2. I extend the timeline from Druckman et al. (2006) (also see Rogowski 2016). The total number of yearly articles in the journal remained fairly constant, and thus any changes in the number of experimental articles do not reflect alterations in the total number of articles published. Finally, as with Druckman et al. (2006), I exclude Gosnell (1926) since he does not employ random assignment or other control mechanisms.

5 Druckman and Green (2021) report results from the same content analysis. However, their coding ended as of May 2019 (and any article posted online at that point), whereas my
the claim that experiments moved from being a nonexistent/marginalized method to an accepted method, and then to a primary method.

While any division into particular “eras of experimentation” would contain an arbitrary element, I label three periods. First is the “preexperimental period” that lasted until roughly 2000. As explained, the end of this period included the publication of some influential experiments; however, applications remained concentrated in a few substantive areas and experiments ostensibly were not a core part of political science curricula. For instance, in their highly impactful 1994 book on social science methodology, King and his colleagues (1994: 125) provide virtually no discussion of experimentation, stating only that experiments are helpful insofar as they “provide a useful model for understanding certain aspects of non-experimental design.”

The “experimental political science 1.0” era, starting around 2000, saw the method become more mainstream/accepted as demarcated by two events. First, in 2000, Gerber and Green (2000) published their field experiment on voter mobilization. This study involved randomly assigning roughly 30,000 registered voters to receive nonpartisan get-out-the-vote

analysis includes all published articles in 2019 (and thus includes nine papers that had not been posted at the time of Druckman and Green’s analysis).

That said, I agree that one could reasonably argue that the 1980s–1990s should be differentiated from the pre-1980s, given the growth of experiments in that time period (see Figure 1.1). Alternatively, one could merge the 1980s–1990s with 2000–2009, although, as explained shortly, two major events (i.e., Gerber and Green (2000) and the start of the Time-Sharing Experiments for the Social Sciences program) signal a qualitative shift in the availability and the use of experiments around 2000.
messages via personal canvassing, direct mail, or telephone calls. One of the more notable findings is that personal canvassing by far has the largest impact on mobilizing voters. The paper offered practical lessons for those interested in increasing turnout and spoke to an ostensible paradox at the time concerning the decline in voting turnout, linking it to the decrease in face-to-face mobilization. The paper accentuated the power of experiments for academics and practitioners. It also was only the third field experiment published in the APSR, and the first in nearly twenty years. The confluence of the widely discussed results along with the reintroduction of the field experimental method stimulated others to turn to field experiments. It sparked burgeoning literature on voter mobilization (e.g., Nickerson 2008; Green et al. 2013) and vote choice (Wantchekon 2003), and more generally, ushered in the use of field experiments in other subfields (e.g., Findley et al. 2014; Hyde and Marinov 2014). It also cohered with the expansion of field experiments in other disciplines (e.g., Bloom 2005).

Second, in 2001, Time-Sharing Experiments for the Social Sciences (TESS) was established with support from the National Science Foundation. TESS capitalizes on economies of scale to enable scholars from across the social sciences, on a competitive basis, to conduct survey experiments on probability-based samples of the US population (see Mutz 2011). Since its founding, TESS has supported more than 600 experiments. Clearly, the first decade of the twenty-first century saw the confirmation of experiments as a mainstream method.

Another change occurred about a decade into the twenty-first century: the establishment of two major experimental institutions. This could be called “experimental political science 2.0.” In 2009, Evidence in Governance and Politics (EGAP) formed as a network for those engaged in field experiments on governance, politics, and institutions. As it grew in membership and capacity, EGAP also expanded its worldwide outreach efforts to include instruction on experimental methods across the Global South. Then, in 2010, the first meeting of the American Political Science Association’s experimental research section took place, and a year later it

---

7 Since 2000, roughly thirty field experiments have been published in the APSR and the Annual Review of Political Science has published several reviews of field experiments on a range of topics, including collective action (de Rooij et al. 2009), developmental economics (Humphreys and Weinstein 2009), political institutions (Grose 2014), and international relations (Hyde 2015).
voted to launch the *Journal of Experimental Political Science* (the first issue of which appeared in 2014).8

These institutions reflected and further stimulated the use of experiments, as is made clear in Figure 1.1 with eighty-four experiments published in APSR since 2010. Not only did the number of experiments substantially increase, but the reach of these experiments expanded to new domains. Examples include the study of elite responsiveness (Butler and Broockman 2011; Grose 2014; Costa 2017); social media communications (e.g., Settle 2018); governmental threats to use force (e.g., Tomz 2007; Kertzer and Brutger 2016); election monitoring (Ichino and Schündeln 2012; Hyde and Marinov 2014; Buzin et al. 2016); foreign aid (Beath et al. 2013); and governance and accountability (Dunning et al. 2019).9 It is telling that, since 2010, roughly 44 percent of the experimental articles published in the APSR can be classified in the field of comparative politics (up from 19 percent during 2000–2009 and 2 percent during 1956–1999).10

In sum, the last decade has seen a dramatic growth of experimental approaches across political science. It is clear that political scientists think about and apply experiments in a very different manner than a decade ago: They think of experimentation as a primary methodology and apply it in novel domains. Understanding this new era of experimentation (starting around 2010) requires more than recognition of the growth of experiments, however. The basic nature of experimentation, across the social sciences, has changed due to technological advances and the open science movement.

**Technological Change and Open Science**

The initial emergence of experiments (in the “preexperimental era”) followed on the heels of several technological advances. In the 1980s,

8 Examples of other institutional developments in political science include the launching of subject pools in more than a dozen departments (Druckman et al. 2018a: 624) and the start of a Routledge book series focused on experimental political science. These institutional innovations were accompanied by some notable publications. This list includes the explosion of experimental articles using Amazon’s Mechanical Turk to furnish research participants (Berinsky et al. 2012; Mullinix et al. 2013) and, in 2011, the *Cambridge Handbook of Experimental Political Science* (Druckman et al. 2011).

9 The last decade also has seen new methods, with experiments increasingly using novel types of designs (e.g., conjoint survey experiments, audits) and samples (e.g., crowdsourcing platforms, social media) (see Druckman and Green 2021 for details). I discuss these developments in later chapters.

10 These percentages come from a content analysis of the articles displayed in Figure 1.1.