

Field Experimentation and the Study of Law and Policy

Donald P. Green¹ and Dane R. Thorley^{1,2}

¹Department of Political Science, ²Columbia Law School, Columbia University, New York, NY 10027; email: dpg2110@columbia.edu, drt2121@columbia.edu

Annu. Rev. Law Soc. Sci. 2014. 10:53–72

The *Annual Review of Law and Social Science* is online at lawsocsci.annualreviews.org

This article's doi:

10.1146/annurev-lawsocsci-110413-030936

Copyright © 2014 by Annual Reviews.
All rights reserved

Keywords

experiments, randomized study, policy evaluation, empirical legal studies, field studies, political science

Abstract

Field experiments are randomized experiments that take place under naturalistic conditions. This research method is experiencing rapid growth throughout the social sciences and especially in legal studies, where it is used to rigorously evaluate policies and programs. We begin by charting the growth of field experimentation in law and legal studies, describing the statistical properties of experiments and discussing the practical threats that may undermine experiments conducted in field settings. Next, we review the experimental research literature in a variety of domains: legal institutions, including the judiciary, legislature, and legal profession; incentives, especially as they apply to tax compliance and business law; and laws and obligations, including legal code, policy, and legal theory. We conclude by highlighting some of the challenges that the experimental literature must confront if it is to speak convincingly to issues of law and policy.

INTRODUCTION

To what extent and in what ways do laws, regulations, rules, and norms affect individuals' behavior and societal outcomes? Could policy makers more efficiently achieve their social, economic, and political objectives by creating, changing, or eliminating institutions or the rules by which they operate? Causal questions such as these draw legal scholars into the realm of social science.

Since the 1970s, a growing number of legal scholars have used or cited quantitative research to address cause-and-effect questions (Diamond & Mueller 2010), and the methodological sophistication of empirical legal studies has grown steadily. In decades past, these scholars relied almost exclusively on regression analysis to extract causal inferences from correlations. For example, Burruss & Kempf-Leonard (2002) sought to test the effects of legal representation on the outcomes for juveniles who faced felony charges in Missouri. In keeping with 10 prior studies, these researchers found that defendants who were represented by attorneys were more likely to be sentenced to a juvenile detention facility or some other form of out-of-home placement. Moreover, this statistical relationship persisted even after the researchers held constant an extensive set of defendant characteristics, such as prior criminal record. Why might representation by an attorney lead to worse outcomes for the defendant? The authors speculate that one reason might be the lack of seriousness with which attorneys defended their juvenile clients, but another explanation may be selection bias: Those who obtained legal counsel may have, from the beginning, been (for reasons that are not fully measured by the researchers) the most vulnerable to an adverse court ruling. Because the researchers did not control access to the treatment, it is unclear whether their results reveal the true adverse effects of legal representation or rather the deficiencies of their research design.

This type of fundamental uncertainty casts a pall over many studies in law and social science. Recognizing that the absence of more telling evidence may bring their research to a deadlock, scholars have gradually changed the way that they approach the study of cause and effect. Prior to what Angrist & Pischke (2010) dubbed the credibility revolution, scholars addressed omitted variables bias as Burruss & Kempf-Leonard (2002) did, by attempting to measure and hold constant confounding variables. Over time, social scientists have become increasingly skeptical of the notion that confounding variables can be enumerated in a systematic, comprehensive, and convincing manner. The growing emphasis on research design rather than post hoc statistical correctives reflects this new orientation. When one studies the effect of a randomly assigned treatment, omitted variable bias is eliminated by design. In comparison with the open-ended set of statistical analyses and robustness checks that accompanies nonexperimental research, a design-driven approach offers a more convincing and transparent form of inquiry. Commenting on the vast literature assessing the effects of legal representation on defendant outcomes, Greiner & Pattanayak (2012, p. 2182) contend that “we know almost nothing as a result of these studies. . . . [T]he only way to produce credible quantitative results on the effect of legal representation is with randomized trials.”

Although the use of experimental methods in law and social science dates back more than a half-century (Farrington & Welsh 2006, Moore & Callahan 1943), experimentation in the legal studies literature has been largely confined to laboratory settings, often using the methods of experimental economics (Roth 1993, 1995). Although there is much to be said for theoretical insights that may be gleaned from laboratory research (Camerer & Talley 2004, Zeiler 2010), the past decade has witnessed the advent of field experimentation as a method to assess causal claims of direct relevance to law and public policy. The term field experiment dates back to early agricultural studies that assessed the effects of shade, water, and fertilizer on crop yield. These “field” studies

were conducted in outdoor farm plots rather than laboratories or greenhouses. In the context of law and social science, field experimentation refers to studies using randomized assignment of treatments that are naturalistic in one or more of the following respects: The treatment used in the study resembles the intervention of interest in the world, the participants resemble the actors who ordinarily encounter these interventions, the context within which subjects receive the treatment resembles the context of interest, and the outcome measures resemble the actual outcomes of theoretical or practical interest (Gerber & Green 2012, pp. 10–11).

Why is naturalism an attractive design feature? Often, those who study law and social science launch experiments to evaluate a new program or potential policy intervention. A study conducted in a naturalistic manner tends to minimize the gap between the causal insight that the experiment furnishes and what the researcher hopes to learn. For example, researchers who seek to study the effects of providing legal assistance to tenants facing eviction may do so by randomly assigning this treatment and assessing outcomes, such as whether a person retains possession of the unit at the end of the litigation (Greiner et al. 2013). Another attractive aspect of field experiment–based research has to do with the unobtrusive manner in which it is often conducted. Social scientists have long questioned the credibility of experiments in which subjects are aware that they are being studied; results may be distorted when people perceive a connection between the treatments they receive and the outcome measures with which they are later presented (Webb et al. 1999). When treatments (e.g., providing an indigent defendant with a public defender) are allocated randomly as part of a day-to-day administrative routine, neither the defendant nor the defender is likely to behave in an artificial manner because they do not perceive themselves to be research participants. Although field experiments are often more difficult and costly to execute than nonexperimental studies, the strengths of this research method have attracted a growing number of scholars in a wide array of disciplines (Gerber & Green 2012, Torgerson & Torgerson 2008).

The recent growth of field experimentation in law and legal studies is evident in recent legal literature. **Figure 1** charts the number of articles in top journals¹ that report the results of an original field experiment or naturally occurring random assignment in a field setting (e.g., random assignment of judges to criminal cases) from 1990 to 2013. Only three such articles were published in these journals in the 1990s; the next did not appear until 2006. Three articles were featured in 2007, and another two appeared in 2010. In 2012 and 2013, eight such articles were published, close to the combined total from all previous years. This pattern of accelerating growth is akin to the early phases of experimental research programs in neighboring disciplines, such as political science (Druckman et al. 2010) and economics (Levitt & List 2009).

Accompanying the growing supply of randomized experiments in field settings has been a renewed demand by those in the legal community for experimentation prior to the development and implementation of laws and policies. Campbell's (1969) early plea for experimentation as a necessary component of policy making and implementation was echoed in Walker's (1988, p. 67) contention that "a program of restricted field experiments should be adopted to predict the impact of proposed changes to the Federal Rules of Civil Procedure." Only recently, however, have legal scholars made the case for experimentation across a broad spectrum of policy domains. McGinnis (2011) argues that the Congressional Budget Office should fund and oversee experimental evaluations before laws are passed. Tingley & Chilton (2013) extend this call for experimentation to the domain of international law, and Ouellette (2013) lays out a broad experimental agenda

¹Publications included the *Journal of Empirical Legal Studies* and the top 10 law journals according to Leiter's (2009) Law School Reports: *Harvard Law Review*, *Yale Law Journal*, *Columbia Law Review*, *Journal of Legal Studies*, *Stanford Law Review*, *University of Chicago Law Review*, *Oxford Journal of Legal Studies*, *Michigan Law Review*, *New York University Law Review*, and *Journal of Law & Economics*.

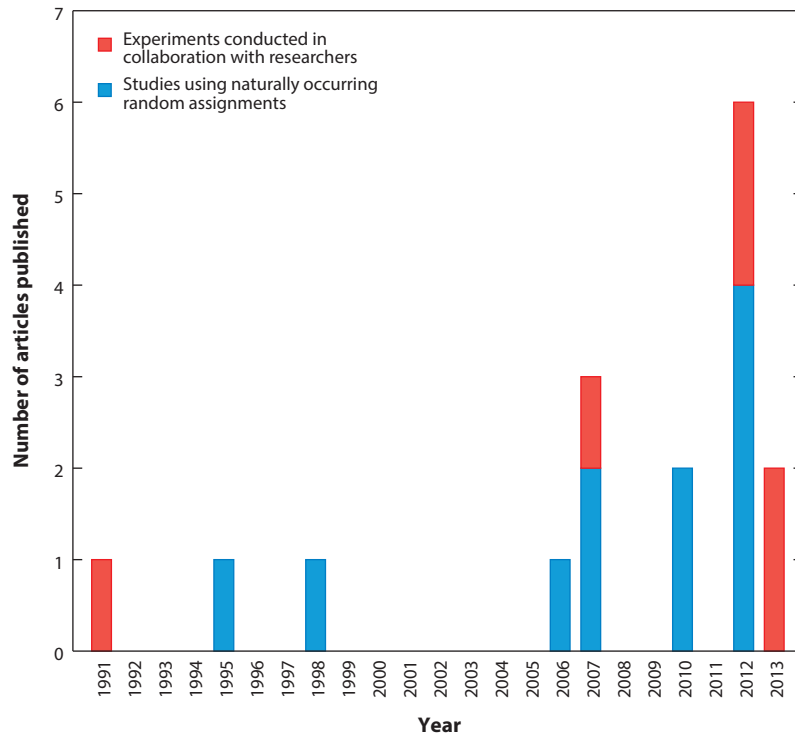


Figure 1

Randomized field experiments published in top law journals. Publications were searched through the HeinOnline database using the following terms: “field experiment,” “randomized,” “naturally occurring randomization,” and “natural experiment.” Excluded are studies that matched these search terms but did not randomly allocate treatments or were situated in something other than a “field” setting. The resulting articles are as follows (in chronological order): Ayres (1991), Ashenfelter et al. (1995), Waldfogel (1998), Harcourt & Ludwig (2006), Abrams & Yoon (2007), Ludwig & Kling (2007), Pleasence & Balmer (2007), Gazal-Ayal & Sulitzeanu-Kenan (2010), Hall (2010), Abrams et al. (2012), Anderson & Heaton (2012), Eigen & Listokin (2012), Eisenberg et al. (2012), Greiner & Pattanayak (2012), Peisakhin (2012), Acquisti et al. (2013), and Greiner et al. (2013).

in the domain of patent law. Abramowicz et al. (2011) go even further, suggesting that experimentation should be used to evaluate the behavioral and societal effects of all aspects of law- and policy making. These arguments reflect a fundamental change in outlook, with a growing number of scholars seeking to place rigorous experimental research at the center of law- and policy making.

Because field experimentation remains a relatively new research method, our review begins with an overview of the formal properties of experiments and the core assumptions that underlie them. We call attention in particular to the ways that practical problems in the field—noncompliance with the assigned treatment and attrition prior to outcome measurement—may affect the interpretation of experimental results. Next, we summarize the experimental research literature in a variety of domains: legal institutions, including the judiciary, legislature, and legal profession; incentives, especially as they apply to tax compliance and business law; and general laws and obligations, including legal code, policy, and legal theory. We conclude by highlighting some of the challenges that confront the continued growth and development of the experimental literature.

ASSUMPTIONS UNDERLYING EXPERIMENTAL DESIGNS

When gauging the effect of an intervention, a researcher must be able to distinguish the treatment effect from incidental differences between treated and untreated subjects. In cases in which the intervention is not randomly assigned, those who receive the treatment may be systematically different from those who do not. This problem is especially acute when subjects self-select into (or out of) treatment. Random assignment can be used to overcome this selection bias. We formalize the problem of estimating treatment effects to underscore the advantages of random assignment and to describe additional assumptions that must be satisfied in order for an experiment to render substantively meaningful estimates. These assumptions have important implications for the scope of feasible field experimentation. We also briefly discuss the issue of generalization beyond the experiment to other people or contexts.

A Potential Outcomes Model of Causal Effects

Suppose the researcher wants to measure the effect of having legal counsel on the probability that a tenant will remain in his or her residence 60 days after the start of an eviction hearing. The researcher gathers information about whether each subject is treated—has legal counsel—or not. We have not yet specified how treatments are assigned; the terminology of treatments and outcomes applies to both experimental studies, in which treatments are assigned at random, and observational studies, in which they are not. We denote the treatment status of subject i by the variable D_i , which takes the value 1 if the subject is treated (has legal counsel) and 0 if not, and we denote the outcome of interest for subject i by Y_i . In this case, we let Y_i equal 1 when an eviction occurs within 60 days and 0 otherwise. In addition to the realized outcome, we may also define a pair of potential outcomes for each subject, which are the hypothetical outcomes that would occur depending on whether the subject is treated or not. We use the notation $Y_i(1)$ to refer to the outcome that a subject would experience if she were treated and $Y_i(0)$ to refer to the outcome that a subject would experience if she were not treated. In the course of conducting an actual study, a researcher will observe only one of these two potential outcomes. The treatment effect for an individual, which in our example is the difference in potential eviction outcomes, is defined as $Y_i(1) - Y_i(0)$. The average treatment effect (ATE) for the collection of subjects is $E[Y_i(1) - Y_i(0)]$, where the $E[\cdot]$ operator refers to the expected or average outcome among all subjects.

Random Assignment as a Remedy to Selection Bias

Regardless of whether a study is experimental or observational, we can calculate the difference in the average values of Y_i for the treated (those with legal counsel) and the untreated (those without):

$$E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 0], \quad 1.$$

where the notation $E[A_i|D_i = B]$ indicates the average value of A_i among those subjects for whom the condition $D_i = B$ holds. What do we learn from comparing average treated outcomes with average untreated outcomes?

Unfortunately, the difference in average eviction rates is not necessarily an unbiased estimator of the average effect of legal counsel. To see how the apparent difference in outcomes can be misleading, rewrite Equation 1 as:

$$E\{[Y_i(1) - Y_i(0)]|D_i = 1\} + E[Y_i(0)|D_i = 1] - E[Y_i(0)|D_i = 0]. \quad 2.$$

The difference in the average outcome of the treated and untreated can be decomposed into the sum of two quantities: The first term is the average treatment effect for a subset of the subjects (the

treated). The latter two terms represent selection bias: the expected difference between what the eviction rate would have been for those who received legal counsel had they not received counsel and what the eviction rate would have been had no one received counsel. Whereas the first term is substantively meaningful and something a researcher might strive to estimate, the selection bias term merely reflects a jumble of unobserved confounders that make the treatment and control groups different in the absence of treatment. The threat of selection bias undermines what an observational study can tell us about the causal effect of a treatment.

Random assignment eliminates selection bias. When treatments are assigned at random, they are by definition independent of subjects' potential outcomes. Under random assignment, the treated potential outcomes of those who receive the treatment are the same, in expectation, as the treated potential outcomes of those who remain untreated:

$$E[Y_i(1)|D_i = 1] = E[Y_i(1)|D_i = 0] = E[Y_i(1)]. \quad 3.$$

By the same reasoning, those randomly assigned to the control group have the same expected untreated potential outcomes as those assigned to receive treatment:

$$E[Y_i(0)|D_i = 0] = E[Y_i(0)|D_i = 1] = E[Y_i(0)]. \quad 4.$$

The implication of Equations 3 and 4 is that when treatments are administered randomly, the selection bias term vanishes and the difference between the treatment and control group means equals the ATE. Substitute Equations 3 and 4 into Equation 2 to obtain:

$$\begin{aligned} & \{E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 1]\} + \{E[Y_i(0)|D_i = 1] - E[Y_i(0)|D_i = 0]\} = \\ & E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 1] = \\ & E[Y(1)] - E[Y(0)] = E[Y(1) - Y(0)] = \text{ATE}. \end{aligned} \quad 5.$$

Random assignment solves the selection problem, which is why it is important to protect the integrity of the random assignment procedure when implementing an experiment. Random allocation of treatments is sometimes subverted by administrators who seek to divert treatments to subjects they regard as especially deserving (Torgerson & Torgerson 2008). For researchers analyzing naturally occurring experiments—for example, the lotteries by which passports are distributed (Clingsmith et al. 2009), soldiers are drafted (Angrist 1990), or seats in selective public schools are allocated (Hastings & Weinstein 2008)—it is imperative to investigate the randomization procedure in order to understand how subjects are allocated. The school choice lotteries studied by Hastings & Weinstein (2008), for example, were quite complex, and the authors painstakingly reconstructed the multilayered lottery before attempting to analyze the results. Some lotteries, on close inspection, turned out not to be lotteries at all. Researchers have analyzed the “random” assignment of appellate court panels (Sunstein et al. 2006), but close inspection of the allocation process revealed that some courts used random assignment, whereas others did not (Hall 2010). Finally, a pristine randomization procedure may be undone by attrition, whereby the outcomes of certain subjects go unmeasured. When attrition is related to potential outcomes, it may introduce bias (Gerber & Green 2012, ch. 7). Sometimes it is difficult to ascertain from the data that are observed whether attrition is related to potential outcomes in a way that would introduce bias. Uncertainty about whether attrition is benign, however, adds an unwelcome layer of uncertainty to the experimental results (Manski 1989, 1995).

Even when researchers are prepared to assume that random assignment was properly implemented and maintained, the assumption that potential outcomes are independent of treatment must be supplemented by two additional assumptions if the researcher is to draw unbiased inferences about the specific causal factors that motivated the experiment. We next discuss these

two important assumptions—exclusion restriction and noninterference—and introduce a note of caution about the extrapolation of results.

Exclusion Restriction

First, the treatment effect must be produced by the treatment itself rather than through some other channel that accompanies a subject's group assignment. To discuss this issue (which is sometimes called internal validity) precisely, we need to introduce some additional notation. Let $Z_i = 1$ if a subject is assigned to the treatment group and $Z_i = 0$ if the subject is assigned to the control group. The exclusion restriction assumption is satisfied when for all $D_i = d$ and $Z_i = z$, $Y_i(d, z) = Y_i(d)$.

This assumption indicates that potential outcomes respond only to the treatment received; the treatment assigned has no additional effect. In other words, a tenant will express the same $Y_i(1)$ so long as she is treated, regardless of whether she was assigned to treatment or control. Unfortunately, this assumption cannot be assessed empirically because we cannot observe what the outcome would have been had this subject been assigned to a different condition.

The exclusion restriction can be jeopardized in two ways. First, the treatment implemented by the experimenter may combine the intervention of interest, D , with some other inadvertent treatment. If so, the experiment provides an unbiased estimate of the combination of the two treatments, not of the specific effect of D . For example, suppose that a legal aid organization provides legal counsel to tenants assigned to the treatment group, and in the course of doing so, the aid organization meets with and advises tenants and their families. The effective treatment then becomes the combination of legal counsel, meetings, and legal aid advice, and the researcher cannot isolate the effect of counsel alone. To link this scenario to the formal statement of the exclusion restriction, notice that $Y(D = 1, Z = 1)$ may not be the same as $Y(D = 1)$ because the assignment to legal counsel entails treatments other than D .

Second, the exclusion restriction may be threatened by other research practices that undermine the symmetry between treatment and control. For example, if the treatment group's legal outcomes are recorded more quickly than the control group's outcomes due to the extra vigilance of the legal aid organization that works with the treatment group, the asymmetry in measurement may bias the researcher's assessment of whether subjects in the treatment group are more likely to be evicted within a certain number of days after the launch of the study.

The basic injunction to experimenters is to avoid doing anything that may produce a difference in the treatment and control group outcomes through a channel other than the intended treatment. The emphasis on unobtrusive designs grows out of concerns about excludability—if subjects perceive the treatment as experimental, any apparent contrast between treatment and control outcomes could be ascribed either to the treatment or to Hawthorne effects (Abramowicz et al. 2011, p. 949). As a practical matter, it is often difficult to create or change laws experimentally without this fact becoming known to the subjects under study. Nevertheless, pilot or demonstration programs have often proven to be highly informative, especially when the effects are so large (or small) that they cannot plausibly be attributed to Hawthorne effects (e.g., Gertler 2004).

The exclusion restriction becomes especially important when experiments fail to administer treatment according to the randomly assigned protocol. It is not uncommon for field experiments to encounter noncompliance, whereby some members of the assigned control group are treated inadvertently or some members of the treatment group fail to receive the treatment. Formally, noncompliance refers to instances in which $Z \neq D$. When an experiment encounters noncompliance, the researcher can no longer obtain unbiased estimates of the average treatment effect among the entire subject pool. Instead, the researcher must settle for consistent estimates of an

average treatment effect among a subgroup. Specifically, the researcher must assume that there are only three kinds of subjects: always-takers, who would receive the treatment regardless of whether they are assigned to treatment or control; never-takers, who would go untreated regardless of their assignment; and compliers, who receive the treatment if and only if they are assigned to the treatment group. Under this assumption about the subject pool and the exclusion restriction, the experiment provides a consistent estimate of the ATE among compliers (Angrist et al. 1996).

To estimate the average causal effect for this subgroup of compliers (the complier average causal effect, or CACE), the researcher first subtracts the eviction rate in the assigned control group from the eviction rate in the assigned treatment group; this estimates the intent-to-treat effect, or the effect of assignment to treatment on outcomes. Next, the researcher subtracts the proportion of subjects in the control group who have legal counsel from the proportion of subjects in the treatment group who have legal counsel. Consistent estimates of the CACE are obtained by dividing the estimated intent-to-treat effect by the difference in treatment rates. For example, suppose 60% of the assigned control group were evicted, as compared with 50% of the assigned treatment group. The intent-to-treat effect would be estimated as $0.5 - 0.6 = -0.1$. Suppose that 30% of the control group obtained legal counsel, as compared with 80% of the treatment group. The estimated CACE would be $-0.1/0.5 = -0.2$, implying that the treatment reduced eviction rates by 20 percentage points among compliers. In sum, noncompliance does not render an experiment useless; rather, it changes the causal question that the experiment addresses and requires the researcher to invoke additional assumptions.

Noninterference

A second key assumption is noninterference across units. This assumption requires that each subject's potential outcomes, $Y_i(1)$ and $Y_i(0)$, are unaffected by the treatments that other subjects are assigned or receive. This assumption is jeopardized when the treatment of one subject has spillover effects on other subjects. For example, if several tenants in the same building are subject to eviction notices, providing one of them with legal counsel may have repercussions for whether others in the building are evicted (e.g., tenants in the treatment group may tell tenants in the control group about strategies their legal counsel used to prevent eviction, which may bias downward the estimated treatment effect). In legal studies, concerns about noninterference are particularly acute when interventions, such as changes in patent laws, cannot be easily confined to a specific geographic region or class of patent (Ouellette 2013). In such cases, treatments have repercussions for units that are nominally in the control group, which may undermine the inferences derived from a simple comparison of treatment and control outcomes. More complex experimental designs are required when spillovers are likely (Hudgens & Halloran 2008).

In sum, although experiments have many attractive properties, their capacity to furnish unbiased causal inferences rests on a set of core assumptions. The assumption that potential outcomes are independent of treatment depends on random assignment, which may be undermined by human tampering or by systematic attrition. The exclusion restriction requires that an experiment manipulate only the treatment of interest while otherwise maintaining symmetry across treatment and control groups. This restriction is especially important when experiments encounter high rates of noncompliance; in this case, even slight violations of the exclusion restriction may lead to substantial bias in the estimation of the average treatment effect among compliers (Angrist & Pischke 2009). Noninterference is also an important design consideration, especially when a treatment administered to one subject may affect other subjects' outcomes due to contagion, communication, or displacement. The threat of interference represents an important practical constraint

on the kinds of experiments that may be carried out in the domain of legal studies, especially when the proposed treatment involves a change in law or legal system (Greiner & Pattanayak 2012).

Extrapolation

Experiments are often used to predict the effects of similar interventions applied to new subjects or in new contexts, or the effects of the same intervention scaled up from a small program to general policy. When the core assumptions discussed above are satisfied, an experiment renders an unbiased estimate of the average treatment effect for particular subjects in a specific context in response to the treatment used. Extrapolation moves us from the relatively firm ground of unbiased experimental procedures to conjectures about what the effects would be of treatments applied to a different set of subjects in other times and places. Unless the experiment includes variation in treatment and context, any extrapolation along these dimensions involves substantial guesswork. Extrapolations become especially tenuous when programs are to be scaled up, which may prompt other actors in the environment to adjust their behavior in ways that materially affect the average treatment effect. For example, if the court system were to require formal legal representation for all tenants facing eviction, the quality of this marginal increment of legal representation might be quite different from the average quality level under experimental conditions. By expressing this note of caution, we do not mean to set an impossibly high standard of evidence or to suggest that methods other than field experimentation are likely to provide more reliable insights; rather, our point is that answering big policy questions requires both thoughtful experimental designs and extensive replication.

FIELD EXPERIMENTS IN LEGAL STUDIES

In this section, we provide an overview of published research in a variety of domains related to law and social science. These research literatures have been grouped according to whether they assess the effects of (a) legal institutions, (b) incentives that might augment or substitute for laws, or (c) laws themselves. Due to space constraints, we exclude from our review the vast experimental research literature on policing and sentencing, which has been summarized in Farrington (2013) and Farrington & Welsh (2006), as well as the literature on promoting collective action and prosocial behavior, which is discussed in de Rooij et al. (2009). Finally, large social experiments, such as Progreso (Gertler 2004), the RAND Health Insurance experiment (Newhouse & Insur. Exp. Group 1993), and Moving to Opportunity (Ludwig et al. 2008), are pertinent to law and policy and illustrate the potential scale that experiments may achieve, but they require an extensive discussion that cannot be included here; the reader is therefore referred to Greenberg & Shroder (2004) and Orr (1999).

Legal Institutions

Legal institutions serve as the juncture between the formation of law and its general implementation. Scholars and policy makers are constantly looking for ways to improve the structure and function of these institutions, an enterprise facilitated by the implementation of randomized experiments.

Legal services. The right to representation in legal proceedings is guaranteed by the US Constitution's Sixth Amendment to anyone facing a loss of liberty (*Argersinger v. Hamlin* 1972). However, what exactly it means to "have the Assistance of Counsel" in these cases has been the topic of extensive debate (Bazelon 1973, Marceau 2012). Even in areas where there is no

constitutional right to representation, such as child custody disputes (Ellis 1990) or welfare termination (Cooper 1980), legal representation may favorably affect the outcomes for those who have the means to hire an attorney.

Scholars have for decades studied the effects of legal services on judicial outcomes in an attempt to guide policy in this area. Some of the earliest examples of this type of empirical legal research are nonexperimental (see Wood 1916), but as computing technology has made possible the analysis of large data sets, researchers have compiled data for thousands of trials, comparing outcomes among those who were represented by legal professionals and those who were not (Nagel 1972). Arguing that this type of nonexperimental comparison is susceptible to selection bias, researchers have since used experimental designs to evaluate the effects of legal representation. Greiner et al. (2013) worked with legal aid programs to randomly assign individuals facing eviction to receive either free attorney representation or unbundled (nonattorney) legal assistance. They found that individuals assigned to work with attorneys were twice as likely to retain possession of their property. These results confirm those reported by Seron et al. (2001), who evaluated the impact of randomized attorney representation in housing disputes and also found positive effects on outcomes. Interestingly, however, similar studies have found that legal assistance had no apparent effect on the probability of retaining unemployment benefits (Greiner & Pattanayak 2012). Recognizing that these seemingly contradictory results may be tied to the specific circumstances studied, these authors and others have called for broad replication of these experiments in other domains (Selbin et al. 2012).

Experiments have also assessed the effects of nontraditional legal assistance. Elbers et al. (2013) measured the impact of electronic legal aid on the physical and mental well-being of personal injury victims by randomly assigning access to a website that provided an e-coach and specific information on the basics of the claims settlement process. The authors found that although the intervention did give participants more realistic expectations regarding the likely compensation amounts, it did not have any significant, long-term effect on their health status, level of depression, ability to work, or sense of empowerment. However, the small apparent intent-to-treat effect of electronic legal aid in part reflects a very low compliance rate. Just one subject in the treatment group actually used the e-coach.

Some courts and agencies in the United States have instituted procedures that randomly assign legal representatives to cases, creating naturally occurring experiments that provide many of the same methodological advantages as field experiments designed by researchers. Abrams & Yoon (2007) used the random assignment of government attorneys to felony cases in Las Vegas to estimate the effects of attorneys' race and professional experience on the severity of sentences that defendants receive. On average, individuals represented by public defenders with 10 years of experience received sentences 17% shorter than those represented by first-year attorneys, and those represented by Hispanic attorneys received sentences that were 26% shorter than those received by defendants represented by black or white public defenders. In a similar study, Anderson & Heaton (2012) analyzed the outcomes of more than 10 years of murder cases in Philadelphia, where one in five indigent murder defendants was randomly assigned public defenders as opposed to appointed counsel. Their results indicate that individuals assigned public defenders had a 19% lower conviction rate, were 62% less likely to receive life sentences, and received 24% less prison time overall. These statistically significant effects suggest that the outcomes of criminal trials are systematically affected by the attributes of the legal representative, which in turn calls into question the reliability of the judicial process and the fairness of unequal representation across socioeconomic lines.

The judiciary and the courts. Holding constant the features of the case before them, to what extent do judges vary in terms of the way they rule? Although skepticism regarding the feasibility

and ethics of randomizing court procedure (Johnson 1985) has traditionally barred researchers from employing randomized treatments in court proceedings, the courts themselves frequently use random assignment for reasons other than scientific investigation (e.g., to avoid political case assignments). These naturally occurring randomizations create opportunities to study the influence of judge characteristics on judicial behavior (Samaha 2009).

Taking advantage of the fact that random assignment determines the composition of US Courts of Appeals panels, Hall (2009) measured how the partisan composition of these panels affects whether or not the US Supreme Court agrees to hear a case and subsequently overturns the decision. He found that from 1994 to 2004, the US Supreme Court was four times more likely to overturn a decision made by an all-Democratic panel than one made by an all-Republican panel. Conversely, Ashenfelter et al. (1995) found that political affiliation had almost no impact on Pennsylvania District Court civil rights cases. Using similar naturally occurring randomizations, scholars have also measured the racial bias of judges in Illinois felony cases (Abrams et al. 2012).

Another way to leverage random assignment of judges is to examine the downstream consequences of varying sentencing philosophies on rates of recidivism. An early study of this kind concluded that drunk-driving defendants who came before a punitive judge were no less likely to reoffend than those who came before a lenient judge (Martin et al. 1993). This finding was echoed by Green & Winik (2010), who studied recidivism among those sentenced in District of Columbia drug courts, and Killias et al. (2000), who studied Swiss convicts randomly assigned to short prison terms or community service.

In addition to exploiting naturally occurring randomizations, researchers have also conducted field experiments to evaluate various adjudicatory practices and policies. Gottfredson et al. (2006) worked with the Baltimore, Maryland, drug courts to test the impact of specialized drug court programs on defendants' drug recidivism rates. Offenders who were randomly assigned to the special programs, which included increased testing and observation, were significantly less likely to be rearrested for similar charges than those who went through normal adjudication and sentencing. Harrell et al. (2000) and Deschenes et al. (1995) executed similar experiments evaluating drug courts in the District of Columbia and Maricopa County, Arizona, respectively. The data from both studies indicated that individuals assigned to the alternative drug court programs subsequently had fewer drug-related charges, although the treated and nontreated individuals in the Maricopa County experiment were equally likely to be arrested for other criminal activities. Likewise, Abrams & Rohlfs (2011) used a 1981 field experiment in Philadelphia (Goldkamp & Gottfredson 1985) in which bail amounts were varied randomly to find the socially optimal bail levels for felony defendants.

Other field experiments have focused on juries and jury behavior. A study by Heuer & Penrod (1989) randomly assigned preliminary and written jury instructions, finding that preliminary instructions significantly increased juror satisfaction and assisted jurors in following correct legal procedure. Contrary to expectations, written instructions seemed to have no impact on the length of deliberation.

Legislatures. Like judicial institutions, legislatures would seem unlikely venues for randomized experiments. Scholars have nevertheless discovered a variety of experimental research opportunities and capitalized on naturally occurring randomizations.

To test the impact of transparency on parliamentary behavior in Uganda, Humphreys & Weinstein (2012) provided information workshops on the performance of the local parliamentary representative in a randomly selected set of legislative districts. Their analysis suggests that although the constituents seemed receptive to information about performance in office, the

transparency treatment had little effect on the legislators' behavior or their prospects for reelection. Scholars have also studied legislative effort using naturally occurring randomization. For example, Titunik (2013) took advantage of the randomized term lengths of senators in Arkansas, Illinois, and Texas to shed light on how legislative behavior is affected by the prospects of an upcoming election. The data showed that senators from Arkansas and Texas who were assigned two-year terms introduced significantly less legislation than senators assigned to four-year terms, whereas shorter term lengths appeared to have no impact on the amount of legislation introduced by senators from Illinois. Dal Bó & Rossi (2011) conducted a similar study in Argentina and found that shorter terms reduced effort and participation among legislators.

Researchers have also used naturally occurring randomizations to study the extent to which legislators use bill sponsorships and committee assignments to enhance their prospects for reelection. Loewen et al. (2013) made use of the fact that the Canadian House of Commons uses a lottery to assign its members the opportunity to propose legislation. They found that, overall, legislators who were randomly assigned the opportunity to propose legislation were not significantly more likely to win reelection; however, lottery-winning legislators who were members of the governing party did have elevated reelection rates. Kellerman & Shepsle (2009) took advantage of a lottery system to study the effect of committee seniority on the career histories of US House of Representatives members; when more than one freshman congressperson is assigned to a committee, the seniority of newly assigned individuals is determined randomly. Lower-ranked members pass fewer sponsored bills, hold lower positions on their original committee, and are more likely to transfer to another committee. However, winning the seniority lottery seems to have no impact on the probability of reelection.

In a few important instances, random assignment has enabled researchers to study the effects of fundamental alterations in electoral rules and deliberation procedures. Working with the Afghan government, Beath et al. (2013) randomly varied the electoral institutions of 250 Afghan villages to compare the impact of at-large elections versus district-bound elections. They found that the at-large system tended to elect politicians who were significantly more likely to have finished high school. Olken (2010) reports the results of an experiment in which Indonesian villages selected local development projects via direct elections or through local councils. Both produced similar substantive outcomes, but direct elections led to substantial increases in villagers' satisfaction with the outcomes.

Finally, a burgeoning research literature examines the reactions of public officials to requests from their constituents, studying in particular whether constituents are treated differently on the basis of their putative ethnicity or party affiliation. Butler & Broockman (2011) sent US state legislators fictitious emails from putatively black or white constituents requesting assistance with voter registration. In some cases, the emails indicated the party with which the constituent intended to register, in order to distinguish racial discrimination from party discrimination. When no party preference was mentioned in the email, Republican legislators were much more likely to reply to the white constituents; the effect was more muted among Democratic legislators, and overall racial discrimination decreased even further when the constituent indicated a party preference. Broockman (2013) used a similar design, contacting state legislators using a putative black alias but varying randomly whether the sender purported to live within or far from each legislator's district. Both black and nonblack legislators were less likely to respond to out-of-state writers, but the effect was significantly smaller among black legislators, suggesting that they had greater intrinsic motivation to help a black person appealing for help. Butler et al. (2012) found representatives to be more responsive to a service request than a policy inquiry or an advocacy statement from a constituent. Whereas the preceding studies looked at legislators' responses to constituents, Broockman (2014) measured the willingness of constituents to contact their

legislators. Constituents from multimember electoral districts in Maryland that had both black and white representatives were invited to communicate with one randomly selected representative. The data show that both black and white constituents were significantly less willing to communicate with representatives of the opposite race when invited to share their opinions on a political issue.

Incentives

Field experiments are a particularly well-suited method to explore the effects of incentives on individual behavior and collective outcomes in ways that may be used to inform law- and policy making.

Tax compliance. Tax evasion is a phenomenon that the government has traditionally addressed through observational studies and trial and error. Experiments, however, allow government agencies to test and refine what are often low-cost methods for increasing tax compliance. Simply sending a letter indicating the probability of an audit significantly increases compliance (Kleven et al. 2011), although the effects are sometimes disappointing when this treatment targets those identified as having a risk of tax evasion (Slemrod et al. 2001). Further research has investigated the efficacy of alternative appeals to pay delinquent taxes. Studying Austrians who were delinquent in paying their TV tax, Fellner et al. (2013) tested the effect of messages that stressed the moral obligation to pay one's fair share and the high rate of tax compliance in the society as a whole, finding that neither normative treatment had a higher impact than a threat of punishment. Castro & Scartascini (2013) conducted a similar experiment in which they included messages appealing to equity, fairness, or deterrence with the tax bills sent to Argentinian property owners, finding that only the deterrence treatment had any significant effect on tax compliance rates. Working with the court system in the United Kingdom, Haynes et al. (2013) tested the impact of sending text messages to individuals with delinquent court fines. Treatments varied from generic messages reminding the individuals of their fines to personalized messages that referred to individuals by name and/or mentioned the specific amounts they owed. Their results showed that text messages increased the rate of payment and messages addressing the recipient by name were the most effective.

Law and business. Closely tied to tax law and policy are the incentive structures related to the incorporation and formalization of businesses. A series of large-scale online field experiments conducted by Baradaran et al. (2013, 2014) showcase the ways in which multiarmed field experiments may shed light on law and policy. Over the course of two years, the researchers sent pseudonymous emails to more than 3,700 incorporation firms requesting assistance in illegally incorporating a company. The emails were embedded with a series of randomly assigned treatments, which varied the country of origin, details regarding the business, and references to legal sanctions and international norms. Whether emails referenced international rules or appealed to norms had little effect on the rate of compliance with the law.

In a study similar to the Baradaran et al. (2013, 2014) experiments, de Andrade et al. (2013) worked with the Brazilian government to find the most cost-effective way to incentivize domestic business owners to legally formalize their companies. Firms were randomly assigned to a control group or one of four treatment groups that either received information on how to formalize; received information, free registration, and the assistance of an accountant; were visited by a municipal enforcement inspector; or had an inspector visit a neighboring firm. Their analysis found that firms formalized only when they were forced to do so through inspector visits, indicating that neither substantial reductions in the cost of registration nor the threat of audit provided sufficient incentives.

Laws and Obligations

Abramowicz et al. (2011, p. 933) recently noted that “experiments have almost never varied the legal rights and obligations of ordinary citizens or entities.” Although it may seem far-fetched to suppose that laws may be varied experimentally, recent studies have proven some ambitious designs to be feasible.

In a study published in 2007, the US Securities and Exchange Commission (SEC) (2007) evaluated a new stock restriction on trading behavior and stock value by randomly assigning this new regulation to a significant portion of stocks in the Russell-3000 equity index. The restriction exempted stocks from short sales. By comparing the treatment group stock prices with those of the control group, which had remained unchanged by the regulation, the SEC was able to evaluate the impact that the short-sale restriction had on market quality and, more broadly, assess the results of regulation on the trading process. One limitation of this study is that it was not unobtrusive, which jeopardizes the exclusion restriction. Because the participants in this study were aware that the regulation in question was being tested, it is unclear whether their behavior was affected by factors other than the intended treatment. Another limitation is the possible violation of noninterference, insofar as treatments applied to some stocks might have repercussions for trades of control stocks.

In addition to evaluating the effects of specific interventions, field experiments have also been used to test the behavioral and theoretical assumptions that underlie laws and code. The Uniform Commercial Code (UCC), a set of recommendations created by prominent legal scholars to harmonize the differing contract laws of individual US states, contains a significant number of default rules meant to reflect the unstated intentions of contracting parties. These rules, which have been widely adopted by various US jurisdictions and international regimes, affect the outcomes of contract disputes but are often based on the intuition and experience of the individuals who have created them. In a field experiment designed to empirically test one of these default rules, Listokin (2010) conducted a series of eBay auctions for used iPods, each of which was randomly assigned one of four warranty types or no warranty at all. One of the warranties matched the default warranty found in the UCC, and Listokin posited that if the UCC default closely approximates precontractual behavior, the iPods with listed warranties matching the default rule would sell for the same amount as the iPods having no warranty at all. The results of this study show that those prices did match, suggesting that the intuition of the framers of the UCC was accurate.

Given the potential of field experiments for testing the impact and efficacy of new laws and for evaluating traditional legal assumptions, we expect increasing use of experimentation in related domains. The rules of civil procedure, for example, have long seemed ripe for experimentation (Walker 1988). Though not a code per se, these rules often influence case outcomes. Seeing the need for evaluation decades before the recent interest in experimentation, the US Court of Appeals measured the impact of pretrial conferences on attorney behavior and the probability of nonadjudicated settlement by randomizing which cases would be subject to the new procedure (Partridge & Lind 1983). The data indicated that the mandatory conferences reduced the number of appeals that reached the judge panels by 10% and increased the overall quality of the briefs that were eventually reviewed, benefits that arguably outweigh the added costs to the parties involved. A far-reaching evaluation of existing rules in comparison to alternative rules and procedures has the potential to increase the efficiency with which vast numbers of disputes are adjudicated.

THE FUTURE OF FIELD EXPERIMENTATION

The rapid growth of field experimentation in law and social science may be credited with sparking the imaginations of researchers around the globe. Pathbreaking experiments have caused

scholars to envision ever more ambitious field studies. As the pace of experimentation accelerates, it becomes easier for researchers to form new research partnerships with governments and NGOs. The result is a rapidly expanding research frontier, aided by developments in the fields of statistics (Angrist et al. 1996, Aronow & Middleton 2013, Lin 2013, Miratrix et al. 2013, Rosenblum & van der Laan 2011) and machine learning (Chipman et al. 2010, Imai & Ratkovic 2013, van der Laan & Rose 2011) that facilitate experimental design and data analysis.

Although attractive in theory, experiments in the domain of law and policy present a number of challenges in practice. To implement an experiment in a real-world setting, researchers must secure and maintain the cooperation of research partners, which may include an array of individuals, organizations, and government entities. As Abramowicz et al. (2011) point out, many governmental organizations are hesitant to implement random assignment. Even when random allocation is uncontroversial, researchers must brace themselves for changes in leadership and priorities that may occur over the course of a study. Because field experiments typically require significant ex ante research design and planning, changes that occur mid-experiment can be costly and disruptive. For this reason, field researchers are often advised to conduct pilot studies in order to identify implementation challenges and to carry out studies at a variety of sites in order to minimize the risk of disruption due to site-specific unforeseen events. Critics of experiments frequently question whether randomized interventions are feasible. Recent decades have shown that scholars are continually finding new and innovative ways to implement random assignment in ways that address increasingly ambitious research questions.

The growing number and sophistication of field experiments have also exposed some important flaws in the intellectual marketplace that vets, publishes, and preserves experimental findings. Throughout the social sciences, there is abundant evidence of publication bias (Gerber & Malhotra 2008, Simonsohn et al. 2014), selective reporting of results (Casey et al. 2012), and deficiencies in the ways in which researchers describe their experimental procedures and findings (Boutron et al. 2008, 2010). Publication bias occurs because journal editors and reviewers favor interesting findings over null findings, a tendency that both distorts the published distribution of experimental findings and creates perverse incentives for authors to selectively report results so as to make a splashier storyline. Leading journals in the biomedical sciences long ago established the requirement that experiments be registered publicly to facilitate comprehensive meta-analyses of both published and unpublished experimental results (De Angelis et al. 2005). They further required that authors adhere to basic reporting standards when presenting experimental designs and results, describing, for example, how the random assignment was performed and whether subjects dropped out of the study before outcomes could be measured (Schulz et al. 2010). Although compliance with these requirements is less than universal (Gill 2012), there is increasing pressure to bring these practices to the social sciences (Miguel et al. 2014) and further to require researchers to specify their statistical analyses in a public registry before obtaining the experimental results, so as to sharpen the distinction between ex ante hypothesis testing and ex post exploratory analysis (Humphreys et al. 2013). Finally, there are increasing calls for postpublication data sharing so that statistical results may be verified by others (Dafoe 2014, Dewald et al. 1986, Evangelou et al. 2005).² Each of these requirements helps maintain the credibility of the experimental enterprise, and these basic elements of scientific infrastructure must be in place before this nascent literature can speak authoritatively to questions of policy.

²Of the 17 articles described in **Figure 1**, only 1 explicitly indicates the availability of replication data (Hall 2010); in the other cases, neither the articles nor the publishing journals' websites provide any guidance on this matter.

Finally, much more scholarly attention must be paid to assessing empirically the generalizability of experimental findings. Speaking credibly to causal questions beyond the immediate scope of any given experiment requires a set of experiments that address plausible conjectures about how treatment effects are likely to vary across subjects, treatments, and contexts. Building this corpus of research requires an extensive and systematic program of follow-up experiments to assess the generalizability of prior results (Cooper et al. 2009) and to test competing hypotheses about the mechanisms through which causal effects operate (Ludwig et al. 2011). Those who aspire to bring credible scientific evidence to bear on the formulation and implementation of laws, policies, and programs should appreciate the enormous and sustained research effort that this undertaking requires.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

The authors thank Alexander Coppock, Daniel Ho, Yair Listokin, and Amber Spry for their helpful comments on an earlier draft.

LITERATURE CITED

- Abramowicz M, Ayres I, Listokin Y. 2011. Randomizing law. *Univ. Pa. Law Rev.* 159(4):929–1005
- Abrams DS, Bertrand M, Mullainathan S. 2012. Do judges vary in their treatment of race? *J. Leg. Stud.* 41(2):347–83
- Abrams DS, Rohlfs C. 2011. Optimal bail and the value of freedom: evidence from the Philadelphia Bail Experiment. *Econ. Inq.* 49(3):750–70
- Abrams DS, Yoon AH. 2007. The luck of the draw: using random case assignment to investigate attorney ability. *Univ. Chic. Law Rev.* 74(4):1145–77
- Acquisti A, John LK, Loewenstein G. 2013. What is privacy worth? *J. Leg. Stud.* 42(2):249–74
- Anderson JM, Heaton P. 2012. How much difference does the lawyer make? The effect of defense counsel on murder case outcomes. *Yale Law J.* 122(1):154–217
- Angrist JD. 1990. Lifetime earnings and the Vietnam era draft lottery: evidence from Social Security administrative records. *Am. Econ. Rev.* 80(3):313–36
- Angrist JD, Imbens GW, Rubin DB. 1996. Identification of causal effects using instrumental variables. *J. Am. Stat. Assoc.* 91(434):444–55
- Angrist JD, Pischke JS. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton Univ. Press
- Angrist JD, Pischke JS. 2010. The credibility revolution in empirical economics: how better research design is taking the con out of econometrics. *J. Econ. Perspect.* 24(2):3–30
- Argersinger v. Hamlin*, 405 U.S. 25 (1972)
- Aronow PM, Middleton J. 2013. A class of unbiased estimators of the average treatment effect in randomized experiments. *J. Causal Inference* 1(1):135–44
- Ashenfelter O, Eisenberg T, Schwab SJ. 1995. Politics and the judiciary: the influence of judicial background on case outcomes. *J. Leg. Stud.* 24(2):257–81
- Ayres I. 1991. Fair driving: gender and race discrimination in retail car negotiations. *Harvard Law Rev.* 104(4):817–72
- Baradaran S, Findley M, Nielson D, Sharman JC. 2013. Does international law matter? *Minn. Law Rev.* 97(3):743–837

- Baradaran S, Findley M, Nielson D, Sharman JC. 2014. Funding terror. *Univ. Pa. Law Rev.* 162(3):477–536
- Bazelon DL. 1973. Defective assistance of counsel. *Univ. Cincinnati Law Rev.* 42(1):1–46
- Beath A, Christia F, Egorov G, Enikolopov R. 2013. *Electoral rules and the quality of politicians: theory and evidence from a field experiment in Afghanistan*. Work. Pap. No. 2013–27, Dep. Polit. Sci., MIT, Cambridge, MA
- Boutron I, John P, Torgerson DJ. 2010. Reporting methodological items in randomized experiments in political science. *Ann. Am. Acad. Polit. Soc. Sci.* 628(1):112–31
- Boutron I, Moher D, Altman DG, Schulz KF, Ravaud P. 2008. Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. *Ann. Intern. Med.* 148(4):295–309
- Broockman DE. 2013. Black politicians are more intrinsically motivated to advance blacks' interests: a field experiment manipulating political incentives. *Am. J. Polit. Sci.* 57(3):521–36
- Broockman DE. 2014. Distorted communication, unequal representation: constituents communicate less to representatives not of their race. *Am. J. Polit. Sci.* 58(2):307–21
- Burruss GW, Kempf-Leonard K. 2002. The questionable advantage of defense counsel in juvenile court. *Justice Q.* 19(1):37–68
- Butler DM, Broockman DE. 2011. Do politicians racially discriminate against constituents? A field experiment on state legislators. *Am. J. Polit. Sci.* 55(3):463–77
- Butler DM, Karpowitz CF, Pope JC. 2012. A field experiment on legislators' home styles: service versus policy. *J. Polit.* 74(2):474–86
- Camerer C, Talley E. 2004. Experimental law and economics. In *Handbook of Law and Economics*, Vol. 2, ed. AM Polinsky, S Shavell, pp. 1619–50. Amsterdam: North Holland
- Campbell DT. 1969. Reforms as experiments. *Am. Psychol.* 24(4):409–29
- Casey K, Glennerster R, Miguel E. 2012. Reshaping institutions: evidence on aid impacts using a preanalysis plan. *Q. J. Econ.* 127(4):1755–812
- Castro L, Scartascini C. 2013. *Tax compliance and enforcement in the Pampas: evidence from a field experiment*. Work. Pap. No. IDB-WP-472, Dep. Res. Chief Econ., Inter-Am. Dev. Bank, Washington, DC
- Chipman HA, George EI, McCulloch RE. 2010. BART: Bayesian additive regression trees. *Ann. Appl. Stat.* 4(1):266–98
- Clingingsmith D, Khwaja AI, Kremer M. 2009. Estimating the impact of the Hajj: religion and tolerance in Islam's global gathering. *Q. J. Econ.* 124(3):1133–70
- Cooper H, Hedges LV, Valentine JC, eds. 2009. *The Handbook of Research Synthesis and Meta-Analysis*. New York: Russell Sage Found.
- Cooper L. 1980. Goldberg's forgotten footnote: Is there a due process right to a hearing prior to the termination of welfare benefits when the only issue raised is a question of law? *Minn. Law Rev.* 64(6):1107–79
- Dafoe A. 2014. Science deserves better: the imperative to share complete replication files. *PS: Polit. Sci. Polit.* 47(1):60–66
- Dal Bó E, Rossi MA. 2011. Term length and the effort of politicians. *Rev. Econ. Stud.* 78(4):1237–63
- de Andrade GH, Bruhn M, McKenzie D. 2013. *A helping hand or the long arm of the law? Experimental evidence on what governments can do to formalize firms*. Work. Pap. No. 6435, Finance Private Sect. Dev. Team, Dev. Res. Group, World Bank, Washington, DC
- De Angelis C, Drazen JM, Frizelle FA, Haug C, Hoey J, et al. 2005. Is this clinical trial fully registered? A statement from the International Committee of Medical Journal Editors. *Ann. Int. Med.* 143(2):146–48
- de Rooij EA, Green DP, Gerber AS. 2009. Field experiments on political behavior and collective action. *Annu. Rev. Polit. Sci.* 12:389–95
- Deschenes EP, Turner S, Greenwood PW. 1995. Drug court or probation? An experimental evaluation of Maricopa County's drug court. *Justice Syst. J.* 18(1):55–73
- Dewald WG, Thursby JG, Anderson RG. 1986. Replication in empirical economics: the Journal of Money, Credit and Banking Project. *Am. Econ. Rev.* 76(4):587–603
- Diamond SS, Mueller P. 2010. Empirical legal scholarship in law reviews. *Annu. Rev. Law Soc. Sci.* 6:581–99
- Druckman JN, Green DP, Kuklinski JH, Lupia A. 2010. Experimentation in political science. In *Cambridge Handbook of Experimental Political Science*, ed. JN Druckman, DP Green, JH Kuklinski, A Lupia, pp. 3–14. Cambridge, UK: Cambridge Univ. Press

- Eigen ZJ, Listokin Y. 2012. Do lawyers really believe their own hype, and should they? A natural experiment. *J. Leg. Stud.* 41(2):239–67
- Eisenberg T, Fisher T, Rosen-Zvi I. 2012. Does the judge matter? Exploiting random assignment on a court of last resort to assess judge and case selection effects. *J. Empir. Leg. Stud.* 9(2):246–90
- Elbers NA, Akkermans AJ, Cuijpers P, Bruinvels DJ. 2013. Effectiveness of a web-based intervention for injured claimants: a randomized controlled trial. *Trials* 14:227
- Ellis JW. 1990. Plans, protections, and professional intervention: innovations in divorce custody reform and the role of legal professionals. *Univ. Mich. J. Law Reform* 24:65–188
- Evangelou E, Trikalinos TA, Ioannidis JPA. 2005. Unavailability of online supplementary scientific information from articles published in major journals. *FASEB J.* 19:1943–44
- Farrington DP. 2013. Longitudinal and experimental research in criminology. *Crime Justice* 42(1):453–527
- Farrington DP, Welsh BC. 2006. A half century of randomized experiments on crime and justice. *Crime Justice* 34(1):55–120
- Fellner G, Sausgruber R, Traxler C. 2013. Testing enforcement strategies in the field: threat, moral appeal and social information. *J. Eur. Econ. Assoc.* 11(2):634–60
- Gazal-Ayal O, Sulitzeanu-Kenan R. 2010. Let my people go: ethnic in-group bias in judicial decisions—evidence from a randomized natural experiment. *J. Empir. Leg. Stud.* 7(3):403–28
- Gerber AS, Green DP. 2012. *Field Experiments: Design Analysis and Interpretation*. New York: Norton
- Gerber AS, Malhotra N. 2008. Do statistical reporting standards affect what is published? Publication bias in two leading political science journals. *Q. J. Polit. Sci.* 3(3):313–26
- Gertler P. 2004. Do conditional cash transfers improve child health? Evidence from PROGRESA's control randomized experiment. *Am. Econ. Rev.* 94(2):336–41
- Gill CJ. 2012. How often do US-based human subjects research studies register on time, and how often do they post their results? A statistical analysis of the ClinicalTrials.gov database. *BMJ Open* 2:e001186
- Goldkamp JS, Gottfredson MR. 1985. *Policy Guidelines for Bail: An Experiment in Court Reform*. Philadelphia: Temple Univ. Press
- Gottfredson DC, Najaka SS, Kearley BW, Rocha CM. 2006. Long-term effects of participation in the Baltimore City Drug Treatment Court: results from an experimental study. *J. Exp. Criminol.* 2(1):67–98
- Green DP, Winik D. 2010. Using random judge assignments to estimate the effects of incarceration and probation on recidivism among drug offenders. *Criminology* 48(2):357–87
- Greenberg D, Shroder M. 2004. *The Digest of Social Experiments*. Washington, DC: Urban Inst. 3rd ed.
- Greiner DJ, Pattanayak CW. 2012. Randomized evaluation in legal assistance: What difference does representation (offer and actual use) make? *Yale Law J.* 121(8):2118–214
- Greiner DJ, Pattanayak CW, Hennessy J. 2013. The limits of unbundled legal assistance: a randomized study in a Massachusetts district court and prospects for the future. *Harvard Law Rev.* 126:901–89
- Hall M. 2009. Experimental justice: random judicial assignment and the partisan process of Supreme Court review. *Am. Polit. Res.* 37(2):195–226
- Hall M. 2010. Randomness reconsidered: modeling random judicial assignment in the U.S. courts of appeals. *J. Empir. Leg. Stud.* 7(3):574–89
- Harcourt BE, Ludwig J. 2006. Broken windows: new evidence from New York City and a five-city social experiment. *Univ. Chic. Law Rev.* 73(1):271–320
- Harrell A, Cavanagh S, Roman J. 2000. *Evaluation of the D.C. Superior Court drug intervention programs*. Res. Brief, Nat. Inst. Justice, Washington, DC
- Hastings JS, Weinstein JM. 2008. Information, school choice, and academic achievement: evidence from two experiments. *Q. J. Econ.* 123(4):1373–414
- Haynes LC, Green DP, Gallagher R, John P, Torgerson DJ. 2013. Collection of delinquent fines: an adaptive randomized trial to assess the effectiveness of alternative text messages. *J. Polit. Anal. Manag.* 32(4):718–30
- Heuer L, Penrod SD. 1989. Instructing jurors: a field experiment with written and preliminary instructions. *Law Hum. Behav.* 13(4):409–30
- Hudgens MG, Halloran ME. 2008. Toward causal inference with interference. *J. Am. Stat. Assoc.* 103(482):832–43
- Humphreys M, Sanchez de la Sierra R, van der Windt P. 2013. Fishing, commitment, and communication: a proposal for comprehensive nonbinding research registration. *Polit. Anal.* 21(1):1–20

- Humphreys M, Weinstein J. 2012. *Policing politicians: citizen empowerment and political accountability in Uganda: preliminary analysis*. Work. Pap., Dep. Polit. Sci., Columbia Univ./Dep. Polit. Sci., Stanford Univ., Stanford, CA
- Imai K, Ratkovic M. 2013. Estimating treatment effect heterogeneity in randomized program evaluation. *Ann. Appl. Stat.* 7(1):443–70
- Johnson SL. 1985. Black innocence and the white jury. *Mich. Law Rev.* 83(7):611–708
- Kellerman M, Shepsle KA. 2009. Congressional careers, committee assignments, and seniority randomization in the U.S. House of Representatives. *Q. J. Polit. Sci.* 4(2):87–101
- Killias M, Aebi M, Ribeaud D. 2000. Does community service rehabilitate better than short-term imprisonment? Results of a controlled experiment. *Howard J. Crim. Justice* 39(1):40–57
- Kleven HJ, Knudsen MB, Kreiner CT, Pedersen S, Saez E. 2011. Unwilling or unable to cheat? Evidence from a randomized tax audit experiment in Denmark. *Econometrica* 79(3):651–92
- Leiter B. 2009. The “top 20” legal journals. *Brian Leiter’s Law School Reports Blog*, Mar. 19
- Levitt SD, List JA. 2009. Field experiments in economics: the past, present, and future. *Eur. Econ. Rev.* 53(1):1–18
- Lin W. 2013. Agnostic notes on regression adjustments to experimental data: reexamining Freedman’s critique. *Ann. Appl. Stat.* 7(1):295–318
- Listokin Y. 2010. The meaning of contractual silence: a field experiment. *J. Leg. Anal.* 3(2):397–416
- Loewen PJ, Koop R, Settle JE, Fowler JH. 2013. A natural experiment in proposal power and electoral success. *Am. J. Polit. Sci.* 58(1):189–96
- Ludwig J, Kling JR. 2007. Is crime contagious? *J. Law Econ.* 50(3):491–518
- Ludwig J, Kling JR, Mullainathan S. 2011. Mechanism experiments and policy evaluations. *J. Econ. Perspect.* 25(3):17–38
- Ludwig J, Liebman JB, Kling JR, Duncan GJ, Katz LF, et al. 2008. What can we learn about neighborhood effects from the Moving to Opportunity experiment? *Am. J. Sociol.* 114(1):144–88
- Manski CF. 1989. Anatomy of the selection problem. *J. Hum. Resour.* 24(3):343–60
- Manski CF. 1995. *Identification Problems in the Social Sciences*. Cambridge, MA: Harvard Univ. Press
- Marceau JF. 2012. Embracing a new era of ineffective assistance of counsel. *Univ. Pa. J. Const. Law* 14(5):1161–218
- Martin SE, Annan S, Forst B. 1993. The special deterrent effects of a jail sanction on first-time drunk drivers: a quasi-experimental study. *Accid. Anal. Prev.* 25(5):561–68
- McGinnis JO. 2011. Laws for learning in an age of acceleration. *William Mary Law Rev.* 53(2):305–50
- Miguel E, Camerer C, Casey K, Cohen J, Esterling KM, et al. 2014. Promoting transparency in social science research. *Science* 343(6166):30–31
- Miratrix L, Sekhon J, Yu B. 2013. Adjusting treatment effect estimates by post-stratification in randomized experiments. *J. R. Stat. Soc. Ser. B* 75(2):369–96
- Moore U, Callahan CC. 1943. Law and learning theory: a study in legal control. *Yale Law J.* 53(1):1–136
- Nagel SS. 1972. Effects of alternative types of counsel on criminal procedure treatment. *Indiana Law J.* 48(3):404–26
- Newhouse JP, Insur. Exp. Group. 1993. *Free for All? Lessons from the RAND Health Insurance Experiment*. Cambridge, MA: Harvard Univ. Press
- Olken BA. 2010. Direct democracy and local public goods: evidence from a field experiment in Indonesia. *Am. Polit. Sci. Rev.* 104(2):243–67
- Orr LL. 1999. *Social Experiments: Evaluating Public Programs with Experimental Methods*. Thousand Oaks, CA: Sage
- Ouellette LL. 2013. *Patent experimentalism*. Work. Pap., Yale Law Sch., New Haven, CT
- Partridge A, Lind A. 1983. *A reevaluation of the Civil Appeals Management Plan*. FJC-R-83-2, Fed. Judic. Cent., Washington, DC
- Peisakhin L. 2012. Transparency and corruption: evidence from India. *J. Law Econ.* 55(1):129–49
- Pleasence P, Balmer NJ. 2007. Changing fortunes: results from a randomized trial of the offer of debt advice in England and Wales. *J. Empir. Leg. Stud.* 4(3):651–73
- Rosenblum M, van der Laan MJ. 2011. Optimizing randomized trial designs to distinguish which subpopulations benefit from treatment. *Biometrika* 98(4):845–60

- Roth AE. 1993. On the early history of experimental economics. *J. Hist. Econ. Thought* 15(2):184–209
- Roth AE. 1995. Introduction to experimental economics. In *The Handbook of Experimental Economics*, ed. JH Kagel, AE Roth, pp. 3–109. Princeton, NJ: Princeton Univ. Press
- Samaha AM. 2009. Randomization in adjudication. *William Mary Law Rev.* 51(1):1–86
- Schulz KF, Altman DG, Moher D, CONSORT Group. 2010. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *Ann. Int. Med.* 152(11):726–32
- SEC (US Secur. Exch. Comm.). 2007. *Economic analysis of the short sale price restrictions under the regulation SHO pilot*. Rep., Off. Econ. Anal., SEC, Washington, DC
- Selbin J, Charn J, Alfieri A, Wizner S. 2012. Service delivery, resource allocation and access to justice: Greiner and Pattanayak and the research imperative. *Yale Law J. Online* 122:45–67
- Seron C, Van Ryzin G, Frankel M, Kovath J. 2001. The impact of legal counsel on outcomes for poor tenants in New York City's housing court: results of a randomized experiment. *Law Soc. Rev.* 35(2):419–32
- Simonsohn U, Nelson LD, Simmons JP. 2014. P-curve: a key to the file drawer. *J. Exp. Psychol. Gen.* 143(2):534–47
- Slemrod J, Blumenthal M, Christian C. 2001. Taxpayer response to an increased probability of audit: evidence from a controlled experiment in Minnesota. *J. Public Econ.* 79(3):455–83
- Sunstein CR, Schkade D, Ellman LM, Sawicki A. 2006. *Are Judges Political? An Empirical Analysis of the Federal Judiciary*. Washington, DC: Brookings Inst.
- Tingley D, Chilton A. 2013. Why the study of international law needs experiments. *Columbia J. Transnatl. Law* 34(1):67–148
- Titunuk R. 2013. *Drawing your senator from a jar: term length and legislative behavior*. Work. Pap., Dep. Polit. Sci., Univ. Mich., Ann Arbor
- Torgerson DJ, Torgerson CJ. 2008. *Designing Randomised Trials in Health, Education and the Social Sciences: An Introduction*. Basingstoke, UK: Palgrave Macmillan
- van der Laan MJ, Rose S. 2011. *Targeted Learning: Causal Inference for Observational and Experimental Data*. New York: Springer
- Waldfogel J. 1998. Reconciling asymmetric information and divergent expectations theories of litigation. *J. Law Econ.* 41(2):451–76
- Walker L. 1988. Perfecting federal civil rules: a proposal for restricted field experiments. *Law Contemp. Probl.* 51(3):67–85
- Webb EJ, Campbell DT, Schwartz RD, Sechrest L. 1999. *Unobtrusive Measures*. Thousand Oaks, CA: Sage
- Wood WJ. 1916. Necessity for public defender established by statistics. *J. Am. Inst. Crim. Law Criminol.* 7(2):230–44
- Zeiler K. 2010. Cautions on the use of economics experiments in law. *J. Inst. Theor. Econ.* 166(1):178–93