ANNUAL REVIEWS

Annual Review of Political Science

Ethics of Field Experiments

Trisha Phillips

Department of Political Science, West Virginia University, Morgantown, West Virginia 26506, USA; email: Trisha.Phillips@mail.wvu.edu

Annu. Rev. Political Sci. 2021. 24:277–300

First published as a Review in Advance on January 13, 2021

The *Annual Review of Political Science* is online at polisci.annualreviews.org

https://doi.org/10.1146/annurev-polisci-041719-101956

Keywords

field experiments, ethics, human subjects, autonomy, consent, institutional review boards, IRBs

Abstract

Political scientists are increasingly conducting field experiments that raise ethical issues that standard review criteria and processes are ill equipped to address. Field experiments can answer important questions, but they can also present various harms to individuals, communities, and political processes; undermine autonomy; introduce partnerships that present complex questions of responsibility; and damage the public's trust in the discipline. This article reviews published empirical and theoretical research, professional guidelines, and media accounts, blog posts, and other sources when appropriate. It characterizes the state of the field regarding the identification of ethical problems, relevant normative guidance, proposed strategies for managing ethical concerns, and issues on the horizon. It concludes that the discipline is making good progress, with robust guidelines and strategies for conducting field experiments ethically. However, there is still much work to be done in refining and expanding current guidance, addressing remaining issues, and promoting norms. This review concludes with some general recommendations for researchers conducting political science field experiments.

1. INTRODUCTION

In the past two decades, political scientists have increasingly turned to field experiments to answer a variety of research questions in different settings and regions. Unlike other research methods, field experiments often involve direct intervention in people's lives, and this can raise new ethical issues that standard review criteria and processes are ill equipped to address. For example, should researchers be concerned about changing the outcome of an election? Should researchers be concerned about randomizing services to people in need? Should researchers be concerned about deceptive studies that recruit subjects by falsely implying the possibility of benefit? And finally, where can researchers turn for guidance on ethical issues related to the design and implementation of field experiments?

This article reviews published empirical and theoretical research on ethics in field experiments, the recently approved *Principles and Guidance for Human Subjects Research* (APSA 2020), and media accounts, blog posts, newsletters, and other sources when appropriate. It presents the state of the conversation as it is happening in print in public forums, and it paints this landscape in broad strokes. Section 2 provides a brief introduction to field experiments. Section 3 presents ethical issues related to harms, benefits, autonomy, partnerships, and professionalism. Section 4 discusses institutional review boards (IRBs), the Belmont Report, and Evidence in Governance and Politics (EGAP) as sources of normative guidance. Section 5 offers additional arguments, proposals, and strategies for managing ethical concerns. Section 6 presents *Principles and Guidance for Human Subjects Research*, recently approved by the American Political Science Association (APSA). Section 7 identifies norm promotion, minimal risk, research exceptionalism, and exploitation as important issues that have been mentioned but have not yet received sustained attention.

This review shows that the discipline is making good progress, with robust guidelines and strategies for conducting field experiments ethically. However, there is still much work to be done in refining and expanding current guidance and strategies, addressing remaining issues, and promoting norms. This review concludes with some general recommendations for researchers conducting political science field experiments.

2. FIELD EXPERIMENTS

Inspired by the success of randomized controlled trials (RCTs) in biomedical research, political scientists are now using RCTs in the field to overcome the limitations of observational studies and laboratory experiments (Gerber 2011). This means that political science researchers working in real-world settings randomly assign subjects to a treatment or control condition and then compare them according to a specified outcome measure. These trials' randomized design and their ability to estimate causal effects make them more appealing than observational field studies, and their real-world setting and potential for generalizability make them more appealing than laboratory experiments. This potential for internal and external validity has made field experiments the "gold standard" in many fields of inquiry (Banerjee & Duflo 2017a).

Field experiments are now common in studies of political behavior, discrimination and corruption, and program evaluation. However, their use is by no means limited to these three topics, and political scientists have used field experiments in just about every field of study. Field experiments share some common features, but can differ in terms of the research question, setting, subject selection, nature of the intervention, outcome measures, and effects on individuals and communities. Researchers can conduct them with or without partners; partners can include governmental agencies, nongovernmental organizations (NGOs), political parties, and campaigns; and the researcher's level of involvement can vary depending on the terms of the partnership.

For a more general discussion of the strengths, weaknesses, and actual and potential uses of field experiments, see Banerjee & Duflo (2017b), Druckman et al. (2011), and Gerber & Green (2012). This review focuses on the ethical issues raised by political science field experiments.

3. IDENTIFICATION OF ETHICAL PROBLEMS

The identification of ethical problems in field experiments comes from a variety of sources, including news media, periodicals, blog posts, newsletters, and commissioned reports, as well as scholarly articles, chapters, and books. Together, these present complaints of actual problems, observations of problematic features of actual experiments (even when there were no reported complaints), and recognition of the potential for problems given common experimental designs and settings. Notably absent is routine reporting of ethical issues identified ex ante and ex post in scholarly presentations of experimental studies.

At this point, there seems to be no common terminology or taxonomy in discussions about the ethical issues associated with field experiments. This review organizes frequently mentioned problems into the following six categories: harms, benefits, risk/benefit ratio, autonomy, partnerships, and professionalism. In identifying problems, this review considers subjects and nonsubjects, individuals and groups, direct and indirect effects, immediate and downstream effects, and effects from the research process and the research outcome (Humphreys 2011, King 2000, Zimmerman 2016). Subjects include treatment and control groups; nonsubjects are individuals and communities who are not enrolled in the research, including enumerators and other research staff, partners, and the profession. Direct effects are those associated with the intervention, whereas indirect effects (also known as collateral effects) are those associated with participation in the research. For example, class credit can be an indirect benefit from participation in some laboratory studies; the cost of parking can be an indirect harm. Immediate effects are (nearly) contemporaneous with the research; downstream effects are time delayed, sometimes significantly. Process effects are those associated with the implementation of the research, and outcome effects are those associated with dissemination of the findings.

3.1. Harms

The list of harms related to field experiments is long and wide ranging. In addition to the rare physical harm, the list includes psychological, social, and economic harms; harms of commission and harms of omission; individual and group harms to subjects and nonsubjects; harms resulting from interference in social and political processes; and outcome harms. This section presents the harms in four categories: interventions, exclusion from interventions, social and political processes, and research outcomes. It concludes with a discussion about overarching issues of normative ambiguity.

3.1.1. Intervention harms. The first category recognizes harms related to the delivery of interventions. These include harms of commission such as direct physical harms [e.g., health-related harms from disconnection of water service in Nairobi slums (Coville et al. 2020)] and direct psychological harms (e.g., threats issued in social shaming mailers, or fear elicited by messages that include identity priming and out-group threat). For example, Morton & Rogers (2016) and Nielson (2016) discuss the anxiety and distress caused by research activities designed to measure religiosity, manipulate religious beliefs, or otherwise require participants to violate cultural or religious norms. This category also includes direct economic harms such as opportunity costs. For example, Gelman (2010) complained about the economic harm he experienced when he was deceived by

a correspondence study and changed his schedule to meet with a (fictional) student. In response to his post, several people shared similar complaints about interruption and loss of productivity. Indeed, many studies nonconsensually consume private or public resources when participation requires the attention of employees who would otherwise be engaged in work-related tasks. As Nathan & White (2021) note, experimenting on street-level bureaucrats can present a risk of "diverting scarce bureaucratic time and resources away from real constituents." In some cases, these opportunity costs can be quantified, and even monetized (Landgrave 2020, Nathan & White 2021, Slough 2018). Slough (2018) timed the phone calls in her study on social welfare programs in Colombia and reports that the intervention consumed 200 hours of bureaucrats' time, with an estimated direct labor cost of $2,644 (US). Gelman (2010) jokingly claims that the deceptive correspondence study consumed $63,000 worth of resources, but Desposato (2021) notes that some deceptive studies treat such large samples that taking even 15 minutes of each subject's time could create an "aggregate harm" equivalent to months of full-time work and large direct labor costs.

This category also includes indirect physical, psychological, social, and economic harms. Desposato (2016b) notes that in some settings, research on political or religious behavior presents a risk of retaliatory violence for subjects and enumerators; Lagunes & Seim (2021) note similar risks for research on corruption. Teele (2014) raises concerns regarding the unintentional physical, psychological, and social harms that can result from social engineering experiments, such as targeted microcredit designed to empower women in low- and middle-income countries. There are also concerns that deceptive research can present risks of psychological and social harms related to inclusion of individuals or groups without their knowledge or consent (see Section 3.4).

3.1.2. Exclusion harms. The second category recognizes harms related to withholding (presumably) beneficial interventions. Sometimes referred to as the "randomization problem" (Baele 2013) or the "ethics of exclusion" (Dionne et al. 2016), this category includes direct harms of omission to subjects in the control groups and indirect harms to other subjects, nonsubjects, communities, and research staff. Subjects in the control group can experience direct harms when a beneficial treatment is withheld from individuals or groups who otherwise would have received it. Examples include information that would inform a choice among political candidates, financial assistance for families at risk of homelessness, cash incentives for behaviors that promote public health, and community access to potable water.

> These studies have resulted in real human costs—some subjects did not receive welfare benefits, some villages did not get a health clinic, and others did not get clean water. Almost certainly someone in a control village in some political science field experiment has died from lack of treatment—although she never consented to have health care or clean water randomized. And while we revel in a glowing reception for our study at a conference, our subjects still do not know that they were subjects or why they did not get a health clinic, they just know that their child is sick and they have nowhere to go. (Desposato 2016b, p. 14)

Subjects in the control group can also experience indirect harms, such as feelings of shame, envy, or injustice when goods are not equally distributed, and other psychological, social, or economic harms. As Carlson (2020, p. 92) notes, "Multiple studies find that cash transfers reduce welfare for control subjects because transfers cause local inflation in the price of food and other necessities."

Subjects in the treatment group (along with enumerators and village chiefs) can also experience retaliatory violence and other psychological and social harms (McDermott & Hatemi 2020). Dionne and colleagues (2016) report several such threats and episodes of violent behavior when conducting research with cash transfers and cash incentives in Malawi.

3.1.3. Social and political harms. The third category recognizes harms related to interventions in social and political processes. These include activities that violate social and legal norms, as well as activities that alter the nature or outcome of political processes. For example, some interventions require research assistants or confederates to violate laws or entice subjects to bribe officials; other interventions stoke racial hostility or inject partisan politics into a nonpartisan election (Johnson 2018, Lagunes & Seim 2021, McDermott & Hatemi 2020, Pan 2021). Interventions designed to change voter behavior can affect whether and for whom an individual votes, and this in turn can affect the distribution of vote share among candidates, the outcome of an election, the behavior of elected officials, and even the distribution of public goods and services (Gubler & Selway 2016, Zimmerman 2016). This category also includes harms related to "foreign intrusion in local politics" (Baele 2013), where researchers from wealthy institutions, with relatively large budgets, intervene in the political processes of foreign countries. The process-related downstream effects of these interventions can create winners and losers and harm individuals and groups. They can also harm entire communities, as was clearly communicated by citizens and public officials in Montana after they learned that researchers had conducted an experiment during an election for state Supreme Court seats (Johnson 2018, Willis 2014).

3.1.4. Outcome harms. This category recognizes harms related to the outcome of the research and dissemination of the research findings. Political or social stigmatization of particular groups or communities are examples of outcome harms. Gubler & Selway (2016) note that when researchers draw subjects from groups rather than a national sample, their findings might not be generalizable, which could in turn harm their subjects. For example, when their own research on caste prejudice sampled only from the lower castes, there was a risk that the upper castes would use their findings as evidence that "caste prejudice was simply a phenomenon of lower castes" (Gubler & Selway 2016, p. 177).

3.1.5. Normative ambiguity. Some of the effects we call harms are normatively ambiguous, meaning that whether a consequence is indeed a harm will depend on the chosen normative theory of analysis (Desposato 2016a). There are at least three sources of normative ambiguity in identifying harms associated with field experiments: conflicting theories of value, descriptive versus normative baselines, and wrongful versus nonwrongful harms.

The lack of an objective theory of value can make it difficult to determine whether some outcomes are harms (McDermott & Hatemi 2020). For example, if experimental interventions encourage subjects to change their childrearing practices or their religious beliefs, whether these outcomes are harms will depend on the comparative value one assigns to particular family structures or religions (Gerber & Green 2012).

A second source of ambiguity is the nature of the baseline. In order to determine whether a consequence is a harm, one needs a baseline against which to measure welfare, and usually the baseline is a common-sensical notion of the normal, natural, or expected course of events (Nozick 1969). However, in difficult cases, two different baselines can emerge: a predicted course of events (a descriptive baseline) and a morally required course of events (a normative baseline) (Nozick 1969). In some cases, actions like withholding a benefit will cause harm relative to one baseline but not the other. For example, if an NGO plans to implement a program that provides school uniforms to children, and the research design introduces randomization, then there are two different baselines one can use to assess harm. The descriptive baseline is one in which all children eligible for the program get school uniforms because this is the predicted course of events; the normative baseline is one in which the children do not get school uniforms because the NGO has no moral obligation to provide them. When measured against the descriptive baseline, schoolchildren are harmed

by the randomization; when measured against the normative baseline, they are not harmed. The choice of baseline, which is similar to what Nickerson & Hyde (2016) refer to as the "relevant counterfactual," is important when determining the ethics of the research (Phillips 2021).

The third source of normative ambiguity comes from the difference between harms that violate a person's rights (wrongful) and harms that do not violate a person's rights (nonwrongful) (Desposato 2021, Feinberg 1984). Moral and legal systems are typically concerned with the former but not the latter; that is, we typically do not think a person does anything morally or legally wrong when they best another person in a competition, speak a painful truth, or deliver a deserved punishment. The idea that public officials might be harmed by the discovery of their corruption or discriminatory practices is not troubling to the average citizen. However, research tends to function according to a different standard in which even nonwrongful harms are to be avoided. For example, it is generally considered important for researchers to maintain strict confidentiality when working with subjects who have broken the law, and it would be wrong for a researcher to play a causal role in the apprehension, conviction, or punishment of a research subject, even if the punishment is deserved. Whether and how this line of reasoning extends to field experiments will determine which nonwrongful harms should be recognized and avoided or minimized.

3.2. Benefits

Field experiments can produce both process- and outcome-related benefits. Many studies implement treatments that provide immediate and downstream benefits to treated subjects and other parties; and many studies produce valuable information about the presence of social and political problems or the effectiveness of strategies to promote social, political, economic, and health-related goods and services. Indeed, researchers are increasingly turning to field experiments because they provide good answers to important questions, and Banerjee, Duflo, and Kremer won the 2019 Nobel Prize in Economic Sciences for using field experiments to address global poverty.

In the context of discussions about ethics, the claim that field experiments provide good answers to important questions is challenged by scholars who raise concerns about the quality of the findings and the importance of the questions. Regarding the quality of the findings, Humphreys (2015) raises concern about post hoc analysis and analytic transparency, and repeats calls for preregistration of experimental design. Carlson (2020), Baele (2013), Desposato (2016a), and Deaton (2020) question the validity, generalizability, and utility of the findings. Carlson (2020) is especially concerned about confounding factors when researchers are testing behavioral theories, or cases in which researchers are not from the area under study and do not know the system or its actors well enough to isolate factors. Baele (2013) and Desposato (2016a) are concerned about the generalizability of findings, especially in program evaluations, where conclusions can be "highly context-dependent and cannot be easily extrapolated from a local tactic to a global strategy against the particular problem" (Baele 2013, p. 13). Deaton (2020) is concerned about the utility of the findings, complicating factors in the implementation process, and the messiness of moving from evidence to policy.

Regarding the importance of the findings, several scholars note that political science does not enjoy the same level of internal and external consensus about value as other disciplines (Carlson 2020, Desposato 2018, Humphreys 2011, Whitfield 2019). For example, while biomedical researchers and the public are united in their belief that health is good and illness is bad, political science "has no similarly strong consensus moral good to which we might uncontroversially direct our research" (Whitfield 2019, p. 530). Desposato (2018) finds that this normative ambiguity in research findings is shared by the public, who do not view all (experimentally generated) knowledge as equally valuable.

3.3. Risk/Benefit Ratio

In many ways, problems in assessing a risk of harm versus potential for benefit are extensions of the problems in identifying harms and benefits. More specifically, overestimating benefits might mean that risks are unjustified. For example, if concerns about validity, generalizability, utility, or normative value cause us to question the benefits associated with research, then the risks of the research might not be justified (Carlson 2020, Desposato 2018, Humphreys 2011). Indeed, Desposato's (2018) public opinion survey showed that the public saw more normative value in studies on discrimination than studies on communication, which meant that even though the studies presented the same amount of risk to subjects, survey respondents considered studies on discrimination to be more acceptable than studies on communication. Given that field experiments often carry more risk of harm than other methods, Baele (2013) and Deaton (2020) argue that the marginal benefits of a field experiment need to be especially clear in order for the marginal risks to be justified.

Similarly, underestimating risks might mean that the benefits are not sufficient to justify the research. For example, if researchers do not fully anticipate or appreciate the psychological or economic harms associated with participation, or the impact of the research on nonsubjects, then the benefits might be insufficient to justify the risks of harm. It is unlikely that researchers anticipated that their deceptive study on discrimination would be so burdensome that subjects like Gelman (2010) would complain and attach a $63,000 cost to the harm caused by the experiment. Risk/benefit ratios need to reflect a comprehensive risk analysis in order for the experiment to be ethical.

Risk/benefit assessments are further complicated by biases and by issues related to commensurability (Findley et al. 2016, Humphreys 2015). Researcher or disciplinary bias could inflate perceptions of benefit, minimize perceptions of risk, or both. Additionally, given the variety of risks and benefits and the diversity of recipients, political science research faces particularly difficult problems of commensurability and relative value. In biomedical research, harms and benefits tend to be measured in terms of recognizable units related to morbidity and mortality, but in many cases of political science field experiments, there are no agreed-upon criteria for assessing the value of the harms and benefits. This lack of clarity makes comparisons difficult and especially vulnerable to researcher bias.

3.4. Autonomy

The fourth type of problem that arises in field experiments encompasses the many ways in which researchers fail to fully respect the autonomy of people who are directly and indirectly engaged by the research. Generally speaking, personal autonomy is an individual's ability to be self-determining, which includes making decisions and acting in accordance with a self-chosen plan (Beauchamp & Childress 2019). Researchers violate autonomy when they enroll subjects without informed and voluntary consent (Teele 2014), or when they inappropriately interfere with a person's decision-making process in other aspects of life. In the context of field experiments, this includes deception and coercion in recruiting practices, as well as manipulation or coercion in treatments.

3.4.1. Recruiting practices. Deceptive recruiting practices include failing to inform (or deliberately misinforming) participants about elements of the research, such as the purpose, the interventions, the risks, or, in some cases, about the research itself. This means that individuals do not have the information they need to choose whether they contribute to certain projects and goals, whether they have certain experiences, or whether they want to be a research subject. Importantly, deception in field experiments can differ from deception in laboratory experiments. In laboratory

experiments, subjects agree to participate in research and know that they are in a research study, though they may be deceived about the nature of the intervention or the purpose of the research. In some field experiments, subjects have not agreed to participate and may never know they were in a research study.

In some cases, the interventions are mild and diffuse (e.g., exposure to a billboard or radio ad); in others, the interventions are more engaging and direct (e.g., individual mailings to home residences or individually addressed emails); and in still others, the interventions have a nontrivial impact on a person's experiences or opportunities. In some cases, the interventions falsely imply the possibility of benefit (e.g., recruiting a graduate student or client) (Desposato 2016a). To be sure, some researchers do obtain informed and voluntary consent from their subjects, and other researchers use modified processes (see Section 5.3 below), but many researchers do not appear to engage in any ex ante or ex post activities designed to respect the autonomy of participants.[1] Reasons cited in the literature include concerns about practicality (e.g., obtaining consent from everyone who might happen to pass a roadside billboard or hear a radio ad), concerns about validity (e.g., social desirability biases and the Hawthorne effect), concerns about standing (from whom would researchers seek consent?), and concerns about research conducted with partners [in some cases withholding consent to the partner intervention "may not be within the rights of the subjects" (Humphreys 2015, p. 104)].

Coercive recruiting practices include withholding a benefit to which a person would otherwise have access, and then offering that benefit in exchange for participation in a study (Phillips 2021). This, of course, raises interesting questions of normative versus descriptive baselines in assessing harm (see Section 3.1, above), but people have a legitimate complaint when a benefit they would have received in the normal course of events is now contingent on participation in a research project. Theoretically, this can happen in any setting, with any type of research, but it is more likely to happen in public policy and program evaluation studies. For example, when New York City's Department of Homeless Services authorized a study to evaluate the effectiveness of a financial assistance program, households at risk of homelessness were offered participation in an RCT (Rolston et al. 2013, p. 3): "After applicants were determined to be eligible for Homebase CP services, they were offered the opportunity to provide informed consent. Participation in the study was voluntary; however, if applicants chose not to consent, they would not be eligible to receive Homebase CP services during the study enrollment period." While the research team claims that consent for this study was voluntary, this claim is debatable because potential subjects would experience harm if they choose not to participate.

3.4.2. Treatments. Beyond recruiting practices, researchers sometimes fail to respect participant autonomy by implementing manipulative or coercive interventions. Some interventions enhance a person's ability to make a rational decision by providing them with information about the availability and features of candidates, services, and courses of action; other interventions undermine a person's ability to make rational decisions by providing false information, provoking fear, or threatening harm. For example, election fliers with false or incomplete information, identity primes and fears of out-group threat, and threats of social shaming are manipulative and coercive and compromise the recipient's ability to engage in rational decision making or act in accordance with their self-chosen plan.

[1] I say "appear" because in many cases we simply do not know whether researchers engage in activities designed to respect the autonomy of participants. It is not common practice among political scientists to discuss consent procedures (or the lack thereof) in scholarly publications and presentation (McDermott & Hatemi 2020).

3.5. Partnerships

Many field experiments are conducted in partnership with government agencies or NGOs, and the terms of these partnerships, the standing of the partner, and the nature of the intervention to be studied can create additional ethical issues for field experiments.

Regarding the terms of the partnership, on the one hand, there is concern that researchers may exploit partners when partners do not have a full understanding of the risks of the research, particularly the risk that the study might produce ambiguous or even unfavorable results (Humphreys 2015). On the other hand, there is concern that generous support from partners could introduce a conflict of interest and compromise the integrity of the research generally.

Another concern is the researcher's degree of agency in or responsibility for the intervention and experimental design (Baele 2013, Humphreys 2015, Nickerson & Hyde 2016). If the research design delivers harm, then it matters whether and to what degree the researcher is responsible. Similarly, if the research design modifies the intervention in a way that introduces harm, then it matters whether and to what degree the researcher is responsible for the modification. Factors that could resolve or exacerbate these problems include the standing of the partner (e.g., human rights organization versus dictatorship) and the legitimacy of the intervention for the partner (e.g., educational programs versus solitary confinement in prisons). These issues, individually and combined, can create significant problems of complicity for researchers conducting field experiments with partners.

3.6. Professionalism

The last category includes the general concern that certain types, features, or consequences of field experiments might erode goodwill and trust between political scientists, partners, and the public (Desposato 2018, Humphreys 2015, Zechmeister 2016). If researchers are not mindful about the ethics of field experiments, and if the discipline does not take seriously the concerns of others, then subjects, host communities, and partners might become unfit or unwilling to work with researchers, and the public might not trust political scientists to act ethically.

Some researchers raise concerns about proper management of subject pools in terms of contamination or pollution (Landgrave 2020, Nathan & White 2021). For example, when discussing the problem of deception, Landgrave (2020, p. 501) notes, "The subject pool of elites is a common resource that we, as a discipline, must learn to share or risk destroying not only for ourselves but future researchers." Beyond fitness to serve as research subjects, individuals, communities, and partners might become unwilling to work with researchers if they no longer trust researchers to treat them ethically (Wilson & Hunter 2010). Public trust is a nebulous good, and it can be difficult to know what it looks like, how to measure it, or indeed, how to tell when it is gone. In some cases, signs that trust is eroding become evident [e.g., public officials becoming annoyed or fatigued by deceptive research (Landgrave 2020)]. In other cases, communities feel the need to protect themselves from researchers and develop their own code of conduct for researchers [e.g., the San people of Southern Africa (Schroeder et al. 2019)]. In yet other cases, the near-complete loss of public trust becomes clear (e.g., African Americans and the healthcare system). The academy is not a nimble creature, and it does not respond quickly to change, so if and when the public loses trust in the acceptability of field experiments, the consequence could be significant and difficult to reverse.

4. SOURCES OF NORMATIVE GUIDANCE

The literature reviewed for this article shows that scholars frequently look to IRBs and the Belmont Report for normative guidance but are becoming increasingly frustrated with their

inability to provide answers on many of the ethical issues raised by field experiments. EGAP also provides principles for "sound and ethical research" (EGAP 2011), but these principles are not referenced frequently and do not provide substantive guidance on many of the issues.

4.1. Institutional Review Boards

Many of the articles reviewed here discuss IRBs and the failure of IRB criteria and processes to ensure that field experiments are ethical. This review assumes that readers have a basic familiarity with the history of human subjects research, the Belmont Report, IRBs, and the US "Common Rule" 45CFR46 [readers who do not should see Morton & Williams (2010) and Yanow & Schwartz-Shea (2016)].

4.1.1. Criteria concerns. Many scholars argue (and one has testified to Congress) that the criteria IRBs use to review research are particularly ill suited for political science research (Hauck 2008). Because the criteria were developed for the purpose of reviewing biomedical research, and then later extended (without revision) to social, behavioral, and economic sciences, the criteria can cause IRB members to see risks where none exist, overlook risks that do exist, and create risk by demanding compliance with IRB rules (Michelson 2016, Yanow & Schwartz-Shea 2016, Zechmeister 2016). For example, Zechmeister (2016) reports that an IRB decided a payment to subjects was coercive even though knowledge of exchange rates and local culture indicated that the payment was equivalent to a can of soda. Driscoll (2016), Michelson (2016), Zechmeister (2016), and Zimmerman (2016) all note that IRB members are not likely to be experts on electoral law or local political, religious, or cultural climates. For example, even if an IRB had reviewed the mailers used in the Montana study mentioned above, it likely would have missed the fact that the fliers violated state law and presented a risk of harm to citizens who were already upset about dark money in local elections. Yanow & Schwartz-Shea (2016) worry that the IRB's often inflexible approach to legal documentation can increase risks for subjects who could have remained anonymous but for the signed consent form.

4.1.2. Process concerns. Many scholars express concern that the IRB process can create a false sense of security about the ethics of the research (Driscoll 2016, Fujii 2012, Michelson 2016, Yanow & Schwartz-Shea 2016, Zechmeister 2016). For example, researchers might think that their research is ethical because the IRB reviewed and approved their study. Or a researcher might think that because a study is classified as exempt from IRB review, it cannot possibly have ethical issues. However, given the shortcomings identified in the criteria used for IRB review, these conclusions are unwarranted. Even worse, this misplaced reliance on the IRB process might discourage the discipline generally, and individual researchers in particular, from considering the ethical aspects of their research (Michelson 2016, Yanow & Schwartz-Shea 2016, Zechmeister 2016). Finally, the often rigid process and categories for IRB review create opportunities and incentives for "work arounds" and "IRB laundering,"[2] (Grimmelman 2015) which exploit a system that was not designed to review research conducted with partners.

[2]According to Grimmelman (2015, p. 255), "IRB laundering" is a strategy through which "academic researchers evade formal ethics-review processes by collaborating with corporate researchers who do experiments and collect data within a company where ethics review processes are looser." In these cases, the researcher does not have any involvement with human subjects, but the researcher is involved in the design of the experiment in ways that will affect human subjects, and the researcher will be involved in the analysis of the data and publication of the study. In some cases, researchers will get approval from an IRB for exempt status on the basis of preexisting data, but critics have noted that when a researcher participates in the design of the study, then it seems disingenuous to claim that the dataset predated the researcher's involvement.

King & Sands (2015, p. 10) recommend that researchers working at institutions with IRBs comply with IRB processes and rules because "the rules exist not only for the protection of human subjects, but also for the protection of the researcher." As they note, universities have an infrastructure designed to protect researchers, but in order to use it, researchers need to cooperate with the appropriate university officials and "make the proposal, get approval, and follow the plan you set out" (King & Sands 2015, p. 4). Additionally, IRBs can sometimes be helpful. According to a recent survey of APSA members, many respondents (96%) considered the IRB to be a good resource for guidance on ethical issues, several respondents (30%) found the IRB feedback to be useful, and some respondents (12%) even reported that their understanding of research ethics was enhanced by the IRB process (Beach & Phillips 2020). However, King & Sands (2015) caution researchers to recognize that IRB approval is "wholly insufficient" for ethical research, and Fujii (2012, p. 718) explicitly states that "the responsibility to act ethically rests ultimately on the individual researcher."

4.2. Belmont Principles

Many scholars look past the IRB process and criteria and focus instead on the principles on which they were based: respect for persons, beneficence, and justice (Natl. Comm. 1979). Indeed, several of the articles included in this review use this normative framework presented in the Belmont Report (see Teele 2014, Glennerster & Powers 2016, McDermott & Hatemi 2020, Zimmerman 2016).

While the Belmont Principles provide a useful heuristic and a good starting point for considering the ethical aspects of field experiments, there is growing dissatisfaction with their ability to identify and resolve the ethical issues in field experiments. The Belmont Report is primarily focused on the protection of the individual subject of research, whereas some political science field experiments (*a*) focus on the behavior of institutions or groups, rather than individual subjects; (*b*) introduce harms and benefits that were neither recognized nor anticipated by the Belmont Report; (*c*) present risks of harm to people who are not the subjects; (*d*) present risks of harm to groups and societies; (*e*) present risks of harm to subjects who may not deserve protection from harms; (*f*) have normatively ambiguous goals; (*g*) are conducted with partners; and/or (*h*) are conducted in settings where there is no expectation of care between the research subjects and the researcher (Humphreys 2015, McDermott & Hatemi 2020, Michelson 2016).

Because the Belmont Principles were not chosen for their ability to guide research that studies power, institutions, programs and policies, and social and political behaviors, the guidance they provide for political science field experiments is incomplete. A straightforward application of these principles might miss important problems or prohibit research that seems to be acceptable on other grounds. For these reasons, there is growing concern that the Belmont Principles are not sufficient as normative guidelines, and when used as a starting point they should be supplemented with theories and principles that address the idiosyncrasies of political science field experiments.

4.3. Evidence in Governance and Politics

In 2011, the Evidence in Governance and Politics[3] network (EGAP) adopted five principles for "sound and ethical practice in the conduct of experimental research on governance and politics

[3]At the time the principles were adopted, the name of the network was Experiments in Governance and Politics.

and in the use of such research for policy and decision-making in the public and private sectors" (EGAP 2011). The first principle, "human subjects protection," confirms that EGAP researchers are committed to ethics in human subjects research, but it does not indicate what ethical treatment requires. In addition to this general commitment to human subjects protection, the principle provides guidance for researchers working with partners. Recognizing that partnerships can happen on a variety of terms, the EGAP principle states that researchers and partners should agree "which party, if either, has primary responsibility for the intervention" (EGAP 2011). Furthermore, "Researchers should disclose the role that they play in the design of interventions implemented by practitioners or third parties." Beyond this, the EGAP principles do not address or provide guidance on any of the problems related to harm, benefit, risk/benefit ratio, autonomy, or professionalism identified in the previous section of this review.

5. ARGUMENTS, STRATEGIES, AND OTHER SOLUTIONS

Beyond IRBs, the Belmont Report, and EGAP, several specific solutions have been proposed to help the discipline and individual researchers navigate these waters. Some are narrow in scope and others broad, some are inchoate and others more fully developed. This section provides a nonexhaustive list of arguments and strategies that seem to be gaining traction in the field.

5.1. Recognizing Communities and Political Processes

Several scholars have argued that field experimentalists need to be more attentive to community harms. Johnson (2018, p. 618) recommends that the discipline work with IRBs to incorporate these considerations into the IRB criteria and processes so that IRBs include "explicit instructions for researchers to consider whether research has the potential to adversely affect a community." However, it is not clear that IRBs want the additional responsibility of implementing review criteria that go beyond their federal mandate, or that IRBs are well positioned to judge whether and to what degree any given study has the potential to adversely affect a community. As noted in the previous section, the ethical evaluation of field experiments is not something that should be outsourced to IRBs. Michelson (2016) suggests that researchers engage community partners and local officials in the planning and implementation of field experiments. When possible and appropriate, this engagement would signal respect and possibly help the researcher identify potential risks to individuals and communities.

Whitfield (2019) and McDermott & Hatemi (2020) argue that it is not sufficient to simply add more criteria to the IRB rubric, or to add a supplemental step of community engagement, but instead field experimentalists need to expand the prevailing normative framework to include principles and values that recognize the community and political values at stake. While the three Belmont Principles of respect for persons, beneficence, and justice provide a good starting point for ethical analysis, "exclusive employment" to these principles leaves political science without the language it needs to critically evaluate the ethical implications of field experiments (Whitfield 2019). That is, a straightforward application of these three principles risks overlooking important ethical concerns, and trying to describe these concerns in the language of the three principles risks losing important features in translation. Whitfield (2019, p. 533) suggests that we add a set of political values to the normative framework that capture "the political autonomy and well-being of individual citizens, the self-determination of communities and certain groups, free association of individuals, and devolution or subsidiarity of some sovereign powers to smaller public and semipublic units."

McDermott & Hatemi (2020) suggest that we add a fourth basic principle, "respect for societies," that requires researchers to consider the potential effects of their interventions on "both local and large scale societal outcomes."

5.2. Identifying and Reducing Harms

Several strategies have been proposed to help the discipline and individual researchers identify, manage, reduce, and report the harms associated with field experiments. These include new research agendas, a renewed focus on cultural sensitivity and situational awareness, a call for ex ante and ex post assessments of harm, and recommendations for sample selection and size.

5.2.1. Empirical ethics. While researchers and others in the discipline struggle to get purchase on these normative issues, Desposato (2018, p. 740) suggests that "empirical ethics" can inform at least some aspects of the debates. According to Desposato, empirical ethics is "asking subjects for their judgements on our research." To be sure, evidence of people's beliefs and opinions cannot answer normative questions (for example, if the majority of people believe that capital punishment is right, that does not make it morally right), but such evidence "can contextualize debates, promote better ethical evaluations, identify unexpected harms, and can shed light on the empirical foundations of ethical questions" (Desposato 2018, pp. 740–41). Rather than field experimentalists hypothesizing about whether an experience would constitute a harm for an abstract individual, empirical ethics simply asks potential subjects to evaluate the experience. This approach could help researchers overcome the difficulties of assuming the perspective of "the concrete other" (Teele 2021), and it could help researchers overcome their own academic and disciplinary biases about the importance of their research. Desposato (2018) and Naurin & Öhberg (2019) have used empirical ethics to explore what political scientists, politicians, and the public think about the appropriateness of deception and other features of GOTV experiments, correspondence experiments, and survey experiments.

5.2.2. Cultural sensitivity. Carlson (2020), Johnson (2018), Morton & Rogers (2016), Michelson (2016), Zechmeister (2016), Pan (2021), and Zimmerman (2016) suggest that, in some cases, researchers could reduce risks of harm to their subjects by being more sensitive to cultural differences in both domestic and international settings (and seeking feedback from area specialists when appropriate). In particular, religious and political differences can render some interventions harmful in one community and not another; even among communities with similar features, different experiences can render some interventions harmful in one community and not another. For example, Morton & Rogers (2016) discuss the importance of recognizing religious opposition to gambling and the wide range of activities that might be construed as gambling; Johnson (2018) and Michelson (2016) discuss the importance of being informed about the local political landscape, in both domestic and international settings. Pan (2021) notes that this "situational awareness" can also help researchers minimize risks to research staff and minimize social impact.

5.2.3. Sample selection and size. Desposato (2016a), Nathan & White (2021), and Slough (2019) suggest that harms related to some experiments can be minimized by ex ante analyses and modifications to research designs. Desposato (2016a) argues that researchers who intervene in political processes (e.g., GOTV experiments) should "tread lightly" by conducting a power analysis and using the smallest possible sample size, choosing an election where polling and history suggest that treatments will not affect the outcome, and not outspending the candidates. Slough (2019) makes similar recommendations and provides a decision rule that helps researchers conduct

an ex ante evaluation of their research design to determine whether the intervention should be implemented or revised.

Nathan & White (2021) make similar recommendations for experiments on street-level bureaucrats. They encourage researchers to make a good faith attempt to quantify the costs of the study, even when the costs might be difficult to calculate, because "even basic and incomplete cost estimates like this are useful, as they force scholars to begin engaging explicitly with the burdens their experimental design may impose." Moreover, in studies

> where adding additional observations is relatively low-cost to researchers, as in an email-based audit experiment, experimenters might simply collect as many observations as is feasible, without considering whether they are "overpowered" for the question at hand. We urge researchers to think about the social costs of additional observations in field experiments just as they might think about budgetary costs in other forms of research, and avoid running experiments that are larger than needed to measure effects of an expected size. (Nathan & White 2021)

Importantly, Desposato (2021), Nathan & White (2021), and Slough (2019) all argue that ethical considerations for identifying and minimizing harm need to be incorporated into the design phase of the research, and all recommend that cost calculations and cost/benefit justification should be reported in preanalysis plans, grant proposals, and published papers.

5.3. Respecting Autonomy: Consent

Many scholars have argued that a renewed focus on consent, along with creatively reimagining the consent process, can address some of the problems associated with field experiments. Alternative models identified and discussed in the literature include implied, deferred, and hypothetical consent (Desposato 2018, Humphreys 2015, Teele 2021). Some of these do a better job than others of respecting participant autonomy, and some address additional problems, but in nearly all cases, employing any one of these methods would be better than ignoring the consent process altogether.

5.3.1. Implied consent. Implied consent, also known as tacit consent, happens when participants are informed about the research project (to some degree) and express consent through their actions. For example, consent is implied when a participant proceeds with a questionnaire after being informed about the research, or when a participant is sent advance notice of a research project and given the opportunity to opt out but fails to do so. Deferred consent occurs when participants are informed of the study only after the research has concluded, and are then offered the opportunity to withdraw from the study. Deferred consent is more robust than debriefing, which simply informs participants that they were involved in a research project. All three models show at least some level of respect for participant autonomy, most are compatible with most types of research, and some could even provide useful information about other ethical issues. For example, as the discipline struggles to identify and assess the significance of harms associated with field experiments, or the degree to which subjects see the research as beneficial, implied and deferred consent (and even debriefing) offer opportunities for subjects to communicate their concerns to researchers. A significant number of subjects opting out, withdrawing, or complaining to the researcher or IRB after debriefing might indicate problems with either the purpose of the research or the perceived risks (Humphreys 2015).

5.3.2. Hypothetical consent. Hypothetical consent, also known as presumed, virtual, or counterfactual consent, is the process through which researchers determine whether a subject would

have consented if asked. This question can be answered through a thought experiment [similar to Rawls's (1971) original position] or it can be answered empirically by asking actual subjects about their hypothetical willingness to participate in certain experiments. Teele (2021) argues that researchers who forego standard and alternative consent processes should, at the very least, engage the hypothetical consent process as a thought experiment. Moreover, when conducting the thought experiment, researchers should assume multiple perspectives in terms of roles, values, and life experiences, and should consider the research project from the perspective of the scientists, subject, community member, and member of a relevant vulnerable, burdened, or marginalized group (Teele 2021). Desposato (2021) makes a similar point when he argues that the appropriateness of an experiment should not be based on only the researcher's perspective because "researchers' own assessments of risk may not accurately reflect the experience of individuals in other contexts."

Desposato (2018) also argues that we should (on occasion) go beyond thought experiments and answer the question empirically. He conducted a study in which he asked subjects (and scholars) to judge the acceptability of two hypothetical field experiments and state their willingness to participate and found

> the most important takeaway is that, for some designs, many subjects would rather not be subjects. . . . These empirical findings should prompt some sober reflection by political scientists. Many of our subjects are placed into our studies against their will. (Desposato 2018, p. 747)

Hypothetical consent explores the space between voluntary, nonvoluntary, and involuntary participation in research. A person participates voluntarily when their preferences are known and support participation; a person participates nonvoluntarily when their preferences are not known; a person participates involuntarily when their preferences are known and opposed to participation. Absent exceptional circumstances (e.g., when a person has no right to refuse a certain intervention), researchers should not conduct deceptive research when they have good reason to believe that subjects would decline if asked.

These alternative consent processes do not resolve all of the problems associated with consent. For example, researchers still need to grapple with issues such as whose consent is needed when studies affect nonsubjects, groups, and communities, and whose consent is not needed when conducting studies on bad actors. However, there is a clear call to consider the "broader menu of possibilities" (Humphreys 2015) both when conducting individual studies and as a separate research agenda to develop new strategies for respecting autonomy in field experiments.

5.4. Partnership Issues: Spheres of Ethics

The EGAP guidelines state that researchers and partners should have an agreement stating "which party, if either, has primary responsibility for the intervention" (EGAP 2011). Additionally, researchers should "disclose the role that they play in the design of interventions implemented by practitioners or third parties" (EGAP 2011). This allocution of responsibility can be useful for researchers seeking to draw on a "spheres of ethics" argument. According to Humphreys (2011), this approach recognizes that there are different standards for different actors, and in a research partnership, the ethical responsibilities can be divided according to the actor's role in the project. The advantage of this proposal is that researchers can study interventions that expose people to a level of risk that is "not normally admissible in research communities" (Humphreys 2011, p. 23). This is important because there are some legitimate groups, with legitimate goals, that implement legitimate interventions that may cause harm to some people. Everybody benefits by knowing whether

these interventions are effective. The disadvantage of this proposal is that its scope is limited by requirements for "partner autonomy" and "partner legitimacy" (Humphreys 2011). In order for spheres of ethics to work, the implementing partner must be autonomous from the researcher. This condition might be easily satisfied when the partner plans an intervention and simply asks the researcher to determine the effectiveness. However, when the researcher also advises on the intervention, then the boundaries of the spheres become blurred. Similarly, in order for spheres of ethics to work, there must be partner legitimacy both in terms of standing and in the intervention. That is, if the partner is seen to be ethically illegitimate in some way, or the intervention is illegitimate even for the partner, then spheres of ethics cannot justify the researcher's involvement in the project.

5.5. Program Evaluations, the Randomization Problem, and the Ethics of Exclusion

When RCTs withhold potentially beneficial interventions from study participants in the control group, there are at least four ways to address this "randomization problem" (Baele 2013) or the "ethics of exclusion" (Dionne et al. 2016).

5.5.1. Scarcity. De La O & Wantchekon (2011) and Banerjee & Duflo (2017a) argue that when implementing agencies do not have the resources to treat all eligible and deserving people, randomizing procedures can provide governments and NGOs with a fair and transparent method for distributing the good or service. Glennerster & Powers (2016) similarly claim that even if some people are denied treatment who otherwise would have received it, as long as the same number of people are treated, there is no harm. Humphreys (2011) argues that this "scarcity argument" is limited in its application because not all goods are scarce (e.g., interventions that provide information have low marginal costs), scarce goods are sometimes divisible, and not all randomizations treat the same number of people. Furthermore, the argument ignores the indirect harms associated with unequal distribution that might render some methods of randomization more ethical than others.

5.5.2. Ignorance. Glennerster & Powers (2016, p. 381) argue that exclusion does not constitute a harm because "we do not know whether the program will have a positive effect so we do not know that any individual. . .is worse off for not receiving the program." Humphreys (2011) argues that this "ignorance argument" is weak because even if partners and researchers do not know for certain whether a treatment is beneficial, the justification for the implementation often provides good reason to believe that it will be beneficial. Furthermore, once a program is determined to be beneficial, this would render all efforts to measure effectiveness or efficiency unethical.

5.5.3. Equipoise. Baele (2013) and MacKay (2018) argue for a more restrictive version of the "ignorance argument" by invoking the biomedical concept of "clinical equipoise," according to which an RCT is ethical only if there is genuine uncertainty in the expert medical community regarding the comparative therapeutic merits of each arm in a trial and the standard of care (if there is one) (Freedman 1987). In short, if there is an effective treatment, then it is unethical to conduct an experiment with a nontreatment control. The principle of clinical equipoise is typically premised on some type of fiduciary duty, and MacKay argues that in some contexts and settings such fiduciary obligations are present in political science field experiments. For example, governmental agencies have a duty to provide basic goods and protect basic human rights, so an RCT program evaluation would be ethical only when policy experts "(1) have no evidence regarding

the relative efficacy of the policy arms; (2) reasonably disagree—given existing evidence—about whether the policy arms are more effective than the BPA [best proven and attainable] policy; or (3) know that they are more effective than the BPA policy" (MacKay 2018, p. 63).

This "principle of policy equipoise" would apply to governmental agencies, groups that have fiduciary or other obligations to the people they serve, and social scientists serving as government-authorized investigators. Baele (2013, pp. 20–21) argues that even when there are no clear fiduciary duties, field experiments in development economics should comply with the principle of equipoise because other participant protections such as risk assessments and rigorous informed consent are often impractical or impossible to achieve.

5.5.4. Justice. Glennerster & Powers (2016) argue that it is acceptable for some individuals to experience harm as long as the burdens for research are fairly distributed. "[G]uaranteeing that no specific individual [coffee farmer] is worse off is not a required or feasible standard for judging ethics; instead the standard is whether coffee farmers in Rwanda in general are or are not likely to be negatively impacted by the research" (Glennerster & Powers 2016, p. 381). Their criteria are similar to the Belmont Report's requirements for justice and the notion that it is acceptable for some individuals to bear the burdens of research that will benefit their group (Natl. Comm. 1979). For example, it would be unjust if individual coffee farmers were made worse off, and coffee farmers as a group were negatively impacted, by research designed to benefit other groups. However, even if coffee farmers as a group are not negatively impacted, and the burdens of participation are distributed in a way that is just, the Belmont Report allows harms only when participation is voluntary and a potential subject could avoid the harms by choosing not to participate in the research. Because exclusion from a program is not voluntary, it is not clear whether group benefit can justify the individual harms in cases like this.

6. APSA's *PRINCIPLES AND GUIDANCE FOR HUMAN SUBJECTS RESEARCH*

On April 4, 2020, APSA approved revisions to its code of ethics, *A Guide to Professional Ethics in Political Science*. The newly adopted *Principles and Guidance for Human Subjects Research* (APSA 2020) (hereafter the *Principles*) articulates 12 principles and guidance for their application to research. The *Principles* is the product of a three-and-a-half-year deliberative process that included formation of an ad hoc committee, support from four different APSA presidents, a membership survey, a public comment period, and approval by the APSA Council.

Having been drafted by political scientists, for political scientists, the *Principles* recognizes that political science research often affects more than the subject, happens in settings where legally documented informed consent is sometimes impractical or inappropriate, has the potential to impact political processes, and is sometimes conducted with partners. The *Principles* reflects this complex nature and recognizes that the complexities may go beyond what can be easily predicted and codified. In almost every case, the *Principles* states a norm, acknowledges that there may be exceptions, and encourages researchers to think about the direct, indirect, intended, and unintended effects of the research process and outcome.

While the *Principles* covers a wide range of disciplinary subfields and settings, and the content is too complex to summarize here, it is worth noting that the document specifically addresses consent, deception, harm, and impact in ways that are relevant to field experiments. The *Principles* does not provide formulas or checklists, nor pretend to present a "one size fits all" solution, but it does state general norms and expectations. For example, the document states that "researchers should generally seek informed consent from individuals who are directly engaged by the research

process," but researchers bear the responsibility to identify the parties from whom they should seek consent, and explain "why consent from these parties was meaningful and sufficient" (APSA 2020, p. 6). In some cases, researchers might follow Michelson's (2016) recommendation and seek consent from local officials and community partners; in other cases, consent from other parties might be meaningful and appropriate. Likewise, the document does not prohibit deception, but does state that researchers "should carefully consider any use of deception" (APSA 2020, p. 7) and encourages researchers to engage other strategies for respecting autonomy such as debriefing, deferred consent, or community review. Importantly, the *Principles* states an expectation that researchers will disclose in publications and presentations whether they used deception and whether they used any alternative strategies for respecting participant autonomy. The *Principles* encourages researchers to think broadly about harms and also to share both an ex ante assessment of harms and an ex post report of harms in their publications and presentations. The document also includes detailed guidance on field experiments that intervene in political processes. Relevant considerations include consent and the impacts on individual experiences, communities, and political outcomes. Any researcher working with a partner is encouraged to be "transparent with the partner about the researcher's objectives and likely risks and benefits arising from the research partnerships" (APSA 2020, p. 16) and fully disclose the nature of the partnership in publications and presentations.

The *Principles* clearly communicates that researchers are responsible for the ethics of their research, and it specifically states that researchers "cannot outsource ethical reflection to review boards, other institutional bodies, or regulatory agencies" (APSA 2020, p. 2). That an IRB might approve a field experiment after review, or approve a field experiment as exempt from review, does not mean that the research is ethical. Ultimately, "researchers have an individual responsibility to consider the ethics of their research related activities" (p. 2). Researchers should consult their own conscience, colleagues inside and outside their immediate circle, colleagues outside their specialty, and the *Principles*. Importantly, the *Principles* also promotes openness and broader discussion by encouraging researchers to disclose specific aspects of their research so that the ethical principles for political science research can evolve to become more reflective of disciplinary values, more informative, and more useful.

7. ISSUES ON THE HORIZON

APSA's *Principles*, the EGAP statement, and the arguments and proposals presented in this review provide robust guidelines and strategies for many of the ethical problems associated with field experiments. However, they do not address all issues, and it is not clear whether or how guidelines and best practices will be promoted. This section lists a few important issues that have not yet received sustained attention.

7.1. Norm Promotion

Several norms have emerged from the literature, EGAP, and APSA's *Principles*. The field now needs to decide how to promote them. Unlike IRBs, codes of ethics and best practices do not have regulatory or institutional enforcement mechanisms, so disciplines and professions must use different tools to foster a culture of research integrity. Common mechanisms include responsible conduct of research (RCR) education and publishing practices, both of which are currently underutilized in the discipline (Baele 2013, Desposato 2016a, McDermott & Hatemi 2020, Nathan & White 2021, Peyton 2013, Zechmeister 2016). Horizontal and vertical integration of formal and informal instruction in ethics can raise awareness of problematic practices, provide students with guidance

and tools, and make professional expectations clear. Journal submission guidelines that clearly state expectations for ethical research and openness can encourage the behavior and further discussion [for example, see the new editorial policy for the *American Political Science Review* (APSR Editors 2021)]. Increasing these and other efforts can promote awareness of and compliance with norms while facilitating the processes of revising, expanding, and developing new norms.

7.2. Defining "Minimal Risk"

The meaning and use of "minimal risk" are well defined in the context of biomedical research but often poorly translated into the context of field experiments. In the biomedical context, minimal risk is a threshold term that serves as a criterion for expedited review and a criterion for consent waivers. That is, if a study presents "no more than minimal risk," then the research does not require full board review, and it might be acceptable for researchers to alter (or forego) the consent process. In the biomedical context, the term means that "the probability and magnitude of harm or discomfort anticipated in the research are not greater in and of themselves than those ordinarily encountered in daily life or during the performance of routine physical or psychological examinations or tests" [45CFR46.102(j)].

In many cases, field experimentalists justify altering or skipping the consent process on the basis that their study presents no more than minimal risk, but this assessment of risk is often based on a transliteration of harms associated with "routine physical and psychological examinations or tests" rather than the social and economic harms that are more common in field experiments. Political science needs to explore, theoretically and empirically, the relevant risks that are "ordinarily encountered," whose "daily life" to use as the standard, and more importantly, the threshold of risk below which it is appropriate to subject a person to research without their consent (Desposato 2016a, Morton & Williams 2010).

7.3. Exceptional Nature of Research

The norms presented in the preceding section place limits on the actions of researchers *qua* researchers that may go beyond restrictions the researcher would face as a private consultant, government or NGO employee, volunteer, or citizen. Some researchers claim that they should be allowed to engage in behaviors that are open to others because researchers are citizens, too. Humphreys (2014) notes that this position "seems to miss the point of research ethics—which is to specify what behavior researchers can expect of each other and the public can expect of researchers beyond what they are prevented from doing by law."

King & Sands (2015) and McDermott & Hatemi (2020) make similar points; the latter argue that "academics have a particular responsibility for protecting the public from unwanted and unknown manipulation." However, it is not clear that their concerns are shared by all. While "research exceptionalism" is an ongoing topic of discussion in other disciplines (Wilson & Hunter 2010), and other sciences are facing a related problem of whether research ethics exists (or should) for citizen scientists (Rasmussen 2021), the issues identified in this review related to roles, responsibilities, and license are unique to political science field experiments and deserve direct attention and discussion.

7.4. Exploitation

The literature shows significant attention to harms and the multiple ways in which individual and group autonomy can be compromised, while giving far less attention to issues of justice and

exploitation. These include (meaningful) return of research findings and, more importantly, the choice of subject population, whether burdens and benefits are fairly distributed, and whether the research is responsive to the needs of the host community (Baele 2013, Teele 2014). For example, Deaton (2020, p. 21) voices concern that RCTs on the welfare system are "not done in the interests of the poor people who were their subjects, but in the interests of rich people (or at least taxpayers or their representatives) who had accepted, sometimes reluctantly, an obligation to prevent the worst of poverty, and wanted to minimize the cost of doing so." Humphreys (2015, p. 87) raises similar concerns with internationally sponsored research, pointing out that "these interventions can sometimes take the form of researchers from wealthy institutions manipulating citizens from poorer populations to answer questions of little interest to those populations."

This can be particularly troubling when researchers choose to conduct research in the developing world because their research would not be acceptable in the United States (Desposato 2016a, McDermott & Hatemi 2020), or because their budget will "go farther" and make them "much more powerful actors than they would be in the [United States]" (Desposato 2016a, p. 267). Other concerns about exploitation can include harmful and covert manipulations of social processes, time-intensive surveys of overburdened populations, and less than minimum-wage payments to MTurk workers [on the last point, see Kwek (2020) and Williamson (2016)]. On these issues, political science might benefit from the discussions of exploitation in other settings (see Phillips 2021) and community-based participatory research as a possible solution (see Israel et al. 2012).

8. CONCLUSION

The articles reviewed herein were written on various topics and in various tones, but many of the early articles focusing on ethics of field experiments were cautionary or even foreboding. They expressed genuine concern for the profession and the people who are directly and indirectly affected by political science field experiments. Humphreys (2011, p. 2), for example, wrote:

> [M]y discussion here should be read as the sometimes pained reflections of a researcher knee deep in these issues rather than as the considered views of a moral philosopher. The view from the trenches is quite grim in that the core questions demand answers but still seem to me largely unanswered and unanswerable.

Authors stated a clear need for the discipline to acknowledge the ethical issues, engage in discussions, and come to a consensus on professional expectations.

The field is still developing, but there are several signs of progress. There is now a discipline-wide recognition of the ethical problems associated with field experiments and a concerted effort to address them in both academic and professional forums, as evidenced by the following three observations. First, most of the articles referenced in this review that address ethics of field experiments were published in the last five years. While they do not constitute a large body of literature, the subfield is growing, and it includes several contributions from early-career scholars. This indicates growing attention to the issues and a welcoming environment for critical dialogue.

Second, consider the contrast between the original and revised editions of the *Cambridge Handbook of Experimental Political Science*. In the first edition, the editors included the following footnote in their introduction (Druckman et al. 2011, FN10):

> Perhaps the most notable topic absent from our introductory chapters is ethics and institutional review boards. We do not include a chapter on ethics because it is our sense that, to date, it has not surfaced as a major issue in political science experimentation.

That volume, which represented the work presented at a conference in 2009, may indeed have expressed the commonly held views of that time. The revised edition (Druckman &Green 2021) contains several chapters that address ethical issues. This indicates that ethics is now embedded in educational resources and discussions of methods.

Third, in many ways APSA's *Principles* answers the call expressed in some earlier pieces. Writing in the *Washington Post*, Humphreys (2014) called for a formal statement of the standards "that we would like the public to expect of us as researchers, and that we, as researchers, would like to be able to expect of each other." While APSA's *Principles*, the EGAP statement, and the other arguments and proposals presented in this review are not definitive solutions to the ethical problems associated with field experiments, they provide robust guidelines and strategies, and they represent a growing awareness, willingness to engage the issues, and material steps forward. One would hope that the early bell ringers would now use a different tone—a tone that expresses excitement about the progress, though reminding readers that there is still much work to be done.

Given the diversity of the field, and the complexity of the ethical issues, it would be inappropriate to provide a checklist or a list of rules here. As Pan (2021) notes, "There are no simple principles that would allow social scientists to make automatic and infallible judgments concerning the ethics of experiments." Readers can find specific recommendations in many of the pieces referenced in this review, and some even include a short list [for example, see Desposato (2021), Humphreys (2014), Nathan & White (2021)]. However, these lists should be read with caution as they are not meant to be exhaustive. Readers can find a more comprehensive list, with detailed guidance, in APSA's *Principles*. Here I conclude with three very general recommendations and echo the emerging chorus of scholars engaging in the conversation:

1. Researchers should take responsibility for the ethical aspects of their research. They cannot simply outsource ethical reflection to IRBs or one-size-fits-all checklists and rules. As King & Sands (2015, p. 7) note, "Getting formal IRB approval is crucial, but wholly insufficient. . .you are the one standing behind your research."
2. Researchers should be informed, thoughtful, and reflexive when considering the ethical aspects of their research. They should be knowledgeable about general and topic-specific ethical practices, as well as issues related to their subject population, setting, intervention, and partnerships (if any). When researchers are working in uncharted waters, "a sometimes useful strategy is to develop and announce some new rules and follow them too, just to make sure and to convey publicly that you are nowhere near the line" (King & Sands 2015, p. 7). Additionally, researchers need to be conscientious and thorough when applying existing guidance, best practices, and rules to their own research.
3. Researchers should be open about the ethical issues they face and the decisions they make. They should explain and justify the ethics of their research in preregistration reports, grant proposals, presentations, and publications. This openness encourages reflection and contributes to the refinement and promotion of norms. Additionally, openness can address reader concerns before they become accusations. This is important for many reasons, one of which is that social media responses can be swift and devastating when researchers appear to ignore the ethical implications of their work.

DISCLOSURE STATEMENT

The author currently serves on the Editorial Board and the Advisory Board for Ethical Research for *American Political Science Review*. The author also served as co-chair of the American Political

Science Association's Ad Hoc Committee on Human Subjects Research, which drafted the *Principles and Guidance for Human Subjects Research*.

ACKNOWLEDGMENTS

I thank Scott Desposato, Macartan Humphreys, Rose McDermott, and Elisabeth Wood, and an anonymous reviewer for their helpful comments and suggestions.

LITERATURE CITED

APSA (Am. Political Sci. Assoc.). 2020. *Principles and Guidance for Human Subjects Research*. **https://connect.apsanet.org/hsr/principles-and-guidance/**

APSR Editors. 2021. Notes from the editors. *Am. Political Sci. Rev.* 115(1):v–viii

Baele SJ. 2013. The ethics of new development economics: is the experimental approach to development economics morally wrong? *J. Philos. Econ.* 7:2–42

Banerjee AV, Duflo E. 2017a. Introduction. See Banerjee & Duflo 2017b, pp. 1–24

Banerjee AV, Duflo E, eds. 2017b. *Handbook of Economic Field Experiments*, Vol. 1. Amsterdam/Oxford: North-Holland

Beach G, Phillips T. 2020. *Survey on research ethics with human subjects: executive summary*. Rep. Ad Hoc Comm. Hum. Subj. Res. **https://www.apsanet.org/Portals/54/goverance/Executive%20Summary%20of%20the%20Survey%20of%20Membership.pdf?ver=2019-08-19-153221-817**

Beauchamp TL, Childress JF. 2019. *Principles of Biomedical Ethics*. New York: Oxford Univ. Press. 8th ed.

Carlson E. 2020. Field experiments and behavioral theories: science and ethics. *PS: Political Sci. Politics* 53:89–93

Coville A, Galiani S, Gertler P, Yoshida S. 2020. *Enforcing payment for water and sanitation services in Nairobi's slums*. NBER Work. Pap. 27569

De La O AL, Wantchekon L. 2011. Experimental research on democracy and development. See Druckman et al. 2011, pp. 384–96

Deaton A. 2020. *Randomization in the tropics revisited: a theme and eleven variations*. NBER Work. Pap. 27600

Desposato S. 2016a. Conclusion and recommendations. See Desposato 2016c, pp. 267–89

Desposato S. 2016b. Introduction. See Desposato 2016c, pp. 1–22

Desposato S, ed. 2016c. *Ethics and Experiments: Problems and Solutions for Social Scientists and Policy Professionals*. New York: Routledge

Desposato S. 2018. Subjects and scholars' views on the ethics of political science field experiments. *Perspect. Politics* 16:739–50

Desposato S. 2021. The ethical challenges of political science field experiments. See Iltis & MacKay 2021. In press

Dionne KY, Harawa A, Honde H. 2016. The ethics of exclusion when experimenting in impoverished settings. See Desposato 2016c, pp. 25–41

Driscoll J. 2016. Prison states and games of chicken. See Desposato 2016c, pp. 81–96

Druckman J, Green DP, eds. 2021. *Advances in Experimental Political Science*. Cambridge, UK: Cambridge Univ. Press. In press

Druckman JN, Green DP, Kuklinski JH, Lupia A, eds. 2011. *Cambridge Handbook of Experimental Political Science*. Cambridge, UK: Cambridge Univ. Press

EGAP (Evid. Governance Politics). 2011. *EGAP Research Principles*. **https://egap.org/wp-content/uploads/2020/05/egap-research-principles.pdf**

Feinberg J. 1984. *Harm to Others: The Moral Limits of the Criminal Law*. New York: Oxford Univ. Press

Findley M, Nielson DL, Desposato S. 2016. Obligated to deceive? Aliases, confederates, and the common rule in international field experiments. See Desposato 2016c, pp. 151–70

Freedman B. 1987. Equipoise and the ethics of clinical research. *N. Engl. J. Med.* 317:141–45

Fujii LA. 2012. Research ethics 101: dilemmas and responsibilities. *PS: Political Sci. Politics* 45:717–23

Gelman A. 2010. $63,000 worth of abusive research. . .or just a really stupid waste of time? *Stat. Model. Causal Inference Soc. Sci. Blog*, May 6. **https://statmodeling.stat.columbia.edu/2010/05/06/63000_worth_of/**

Gerber AS. 2011. Field experiments in political science. See Druckman et al. 2011, pp. 115–38
Gerber AS, Green DP. 2012. *Field Experiments: Design, Analysis, and Interpretation*. New York: W.W. Norton
Glennerster R, Powers S. 2016. Balancing risk and benefit: ethical tradeoffs in running randomized evaluations. In *The Oxford Handbook of Professional Economic Ethics*, ed. GF DeMartino, D McCloskey. New York: Oxford Univ. Press
Grimmelmann J. 2015. The law and ethics of experiments on social media users. *Colo. Technol. Law J.* 13:219–72
Gubler J, Selway J. 2016. Considering the political consequences of comparative politics experiments. See Desposato 2016c, pp. 171–82
Hauck RJP. 2008. Protecting human research participants, IRBs, and political science redux: editor's introduction. *PS: Political Sci. Politics* 41:475–76
Humphreys M. 2011. Ethical challenges of embedded experimentation. *APSA Comp. Democr.* 9:10, 23–29
Humphreys M. 2014. How to make field experiments more ethical. *Washington Post Monkey Cage Blog*, Nov. 17. **https://www.washingtonpost.com/news/monkey-cage/wp/2014/11/02/how-to-make-field-experiments-more-ethical/**
Humphreys M. 2015. Reflections on the ethics of social experimentation. *J. Glob. Dev.* 6:87–112
Iltis A, MacKay DP. 2021. *Oxford Handbook of Research Ethics*. Oxford Univ. Press. In press
Israel BA, Eng E, Schulz AJ, Parker EA. 2012. *Methods for Community-Based Participatory Research for Health*. San Francisco: Wiley
Johnson JB. 2018. Protecting the community: lessons from the Montana Flyer Project. *PS: Political Sci. Politics* 51:615–19
King G, Sands M. 2015. *How human subjects research rules mislead you and your university, and what to do about it*. Work. Pap., Harvard Univ., Cambridge, MA. **https://gking.harvard.edu/files/gking/files/irb_politics_paper_1.pdf**
King N. 2000. Defining and describing benefit appropriately in clinical trials. *J. Law Med. Ethics* 28:332–43
Kwek A. 2020. Crowdsourced research: vulnerability, autonomy, and exploitation. *Ethics Hum. Res.* 42:22–35
Lagunes P, Seim B. 2021. The state of experimental research on corruption control. See Druckman & Green 2021. In press
Landgrave M. 2020. Can we reduce deception in elite field experiments? Evidence from a field experiment with state legislative offices. *State Politics Policy Q.* 20:489–507
MacKay DP. 2018. The ethics of public policy RCTs: the principle of policy equipoise. *Bioethics* 32:59–67
McDermott R, Hatemi PK. 2020. Ethics in field experimentation: a call to establish new standards to protect the public from unwanted manipulation and real harms. *PNAS* 117:30014–21
Michelson MR. 2016. The risk of over-reliance on the Institutional Review Board: an approved project is not always an ethical project. *PS: Political Sci. Politics* 49:299–303
Morton RB, Rogers J. 2016. Religion, experiments, and ethical concerns. See Desposato 2016c, pp. 66–80
Morton RB, Williams KC. 2010. *Experimental Political Science and the Study of Causality: From Nature to the Lab*. Cambridge, UK: Cambridge Univ. Press
Nathan NL, White A. 2021. Experiments on and with street-level bureaucrats. See Druckman & Green 2021. In press
Natl. Comm. Prot. Hum. Subj. Biomed. Behav. Res. 1979. *The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research*. Washington, DC: Gov. Print. Off.
Naurin E, Öhberg P. 2019. Ethics in elite experiments: a perspective of officials and voters. *Br. J. Political Sci.* **https://doi.org/10.1017/S0007123418000583**
Nickerson DW, Hyde S. 2016. Conducting research with NGOs: relevant counterfactuals from the perspective of subjects. See Desposato 2016c, pp. 198–216
Nielson RA. 2016. Ethics for experimental manipulation of religion. See Desposato 2016c, pp. 42–65
Nozick R. 1969. Coercion. In *Philosophy, Science, and Method: Essays in Honor of Ernest Nagel*, ed. S Morgenbesser, P Suppes, M White, pp. 440–72. New York: St. Martin's
Pan J. 2021. Experiments on political activity governments want to keep hidden. See Druckman & Green 2021. In press
Peyton K. 2013. Ethics and politics in field experiments. *Exp. Political Sci.* 3:20–36
Phillips T. 2021. Coercion, exploitation, and undue inducement. See Iltis & MacKay 2021. In press

Rasmussen LM. 2021. Research ethics in citizen science. See Iltis & MacKay 2021. In press
Rawls J. 1971. *A Theory of Justice*. Cambridge, MA: Harvard Univ. Press
Rolston H, Geyer J, Locke G. 2013. *Final report: evaluation of the Homebase Community Prevention Program*. Rep., NYC Dep. Homeless Serv., New York, NY. **https://www.abtassociates.com/insights/publications/report/evaluation-of-the-homebase-community-prevention-program-final-report**
Schroeder D, Chatfield K, Singh M, Chennells R, Herissone-Kelly P. 2019. The San code of research ethics. In *Equitable Research Partnerships: A Global Code of Conduct to Counter Ethics Dumping*, ed. D Schroeder, K Chatfield, M Singh, R Chennells, P Herissone-Kelly, pp. 73–87. Cham: Springer
Slough T. 2018. *Bureaucrats driving inequality in access: experimental evidence from Colombia*. Work. Pap., Columbia Univ., New York, NY. **http://taraslough.com/assets/pdf/JMP.pdf**
Slough T. 2019. *The ethics of electoral experimentation: design-based recommendations*. Work. Pap., Columbia Univ., New York, NY. **http://taraslough.com/assets/pdf/eee.pdf**
Teele DL. 2014. Reflections on the ethics of field experiments. In *Field Experiments and Their Critics: Essays on the Uses and Abuses of Experimentation in the Social Sciences*, ed. DL Teele, pp. 115–40. New Haven/London: Yale Univ. Press
Teele DL. 2021. Virtual consent: the bronze standard for experimental ethics. See Druckman & Green 2021. In press
Whitfield G. 2019. TRENDS: Toward a separate ethics of political field experiments. *Political Res. Q.* 72:527–38
Williamson V. 2016. On the ethics of crowdsourced research. *PS: Political Sci. Politics* 49:77–81
Willis D. 2014. Professors' research project stirs political outrage in Montana. *NY Times*, Nov. 17. **https://www.nytimes.com/2014/10/29/upshot/professors-research-project-stirs-political-outrage-in-montana.html**
Wilson J, Hunter D. 2010. Research exceptionalism. *Am. J. Bioethics* 10:45–54
Yanow D, Schwartz-Shea P. 2016. Encountering your IRB 2.0: what political scientists need to know. *PS: Political Sci. Politics* 49:277–86
Zechmeister EJ. 2016. Ethics and research in political science: the responsibilities of the researcher and the profession. See Desposato 2016c, pp. 255–61
Zimmerman B. 2016. Information and power: ethical considerations of political information experiments. See Desposato 2016c, pp. 183–97