

Subjects and Scholars' Views on the Ethics of Political Science Field Experiments

Scott Desposato

Recent controversies raise questions regarding the ethics of political science field experiments. I present here results from a public opinion survey in which subjects and scholars evaluated the acceptability of two hypothetical field experiments. In the survey, the designs were randomly varied to identify the most controversial features. Both scholars and subjects reacted negatively to deception and to experiments without informed consent, especially when the research aims were normatively ambiguous. In some cases, half of the respondents reported that they would rather not be in a typical field experiment without their consent.

Twenty-five years ago, political science experiments were relatively rare, restricted to a handful of subfields, and typically involved a few undergraduates playing games or watching videos in unused classrooms. Today, political science scholars across all subfields are using experimental methods in almost every country, sometimes with thousands of subjects.¹

This dramatic growth in experimental political science has been accompanied by new ethical issues. Many of the issues involve field experiments conducted without informed consent and embedded in political processes, including elections. In some, scholars send political information to voters, including simple turnout reminders, information about polling places, or advertisements criticizing candidates. In other designs, scholars interact with subjects and pretend to be some third party, making some informational request. In both cases,

subjects do not consent to participate and typically never know that they are participating in research.

Such studies are allowable under the Common Rule and most have institutional review board (IRB) approval.² Yet they seemingly clash with ethical norms of voluntary participation in research. In addition, when detected by the public, these clandestine studies often generate controversy and anger, suggesting that the subjects in such studies do not always want to be subjects. In one example, a flyer with information about candidates in Montana led to public anger, accusations of election tampering, and apologies from the presidents of the universities involved in the study.³ This case and others raise questions of involuntary participation in research and potential harms of research on political processes. The controversies observed also suggest that such studies may erode public trust in and support for research.

The issues raised by these types of studies are important, and need more than theoretical ethics debate. A critical missing piece of the puzzle is how subjects feel about their participation in our research. They are included without their consent in many studies and usually never know that they are participating in research. Controversies and anecdotes suggest that at least some would prefer to be excluded from field experiments, but we do not know the extent of public disapproval of such studies and the precise features of field experiments that may upset subjects.⁴

I seek to contribute to our understanding of these issues through a public opinion survey on the ethics of political science field experiments. I focus on correspondence studies and informational field experiments. Respondents evaluated two standard designs and reported their opinions. Subjects came from a sample of adult

*Data replication sets are available in Harvard Dataverse at: <https://doi.org/10.7910/DVN/1WDVZB>

A list of permanent links to Supplementary Materials provided by the authors precedes the References section

Scott Desposato is Professor of Political Science at University of California, San Diego (swd@ucsd.edu). He thanks Seth Hill, Seonghui Lee, Steve Smith, Marco Steenberg, Dawn Teele, Liz Zechmeister, the Institut für Politikwissenschaft, Universität Zurich, and the American Political Science Association. This research was approved by the University of California, San Diego Human Research Protections Program. Analysis and design were pre-registered at Open Science.

doi:10.1017/S1537592717004297

© American Political Science Association 2018. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

residents of the United States. I compare subjects' opinions with those of scholars who were surveyed, using a separate sample of APSA membership.

I preview two findings. First, field experiments without consent were consistently viewed as less acceptable than those conducted with the consent of subjects. In some designs up to half of the subjects reported that they would rather not be included in a study without their expressed consent. Second, opinions varied with the nature of the study. Deceptive research with clear public benefit was judged more acceptable than research with more ambiguous benefit.

Political Science Field Experiments and New Ethical Challenges

I focus herein on two types of field experiments that clearly fall within the boundaries of political science, are typical of work in that field, and where political scientists are often entirely responsible for the study: informational field experiments (IFE) and correspondence study field experiments (CSFEs).⁵

With IFEs, researchers provide subjects with information, then observe behavior. For example, scholars might send information about an election to subjects, then observe whether or for whom subjects vote. In CSFEs, researchers interact with subjects, pretending to be some third party. Researchers then measure whether or how the subject responds. For example, scholars might contact politicians, pretending to be constituents asking for assistance. In other contexts, scholars might pretend to be activists and ask members of the public to sign a petition or to take other action. Closer to home, scholars might pretend to be students and e-mail faculty asking for data.

These designs offer many potential advantages. They avoid Hawthorne effects, as subjects do not know that they are being studied. Field experiments also provide potentially more generalizable and policy-useful information about causal effects, as research is conducted in natural settings with populations of interest rather than in laboratories with convenience samples of college students.⁶

However, these studies also pose two broad sets of related ethical challenges, involving potential harms and the lack of informed consent of subjects. Regarding the first issue, there are many potential effects of field experiments, but it is unclear which should be considered harm. For example, most IFEs and CSFEs pose only trivial *physical* risk to subjects. For example, the physical risk to an individual subject of a study involving a flyer during a campaign might only be that of a paper cut.⁷ However, such studies may have many other potential negative effects on individual subjects. One is emotional harm; recipients of typical turnout or "Get Out the Vote" (GOTV) social pressure letters have reported feeling

offended, shamed, and harrassed.⁸ Subjects in CSFEs are sometimes upset and complain about wasted time.⁹ Other subjects have even filed lawsuits against scholars when they detected the study.¹⁰

Besides these individual harms, field experiments may have broader aggregate or social impacts that may be considered harm.¹¹ Political processes, especially elections, naturally aggregate many small decisions and actions into larger effects. An IFE might significantly change vote share or even an election outcome. A CSFE of politicians might change patterns of representation and cause politicians to reallocate time from constituency service to bureaucratic tasks. For some, such impacts on political processes are harmful, because of the zero-sum nature of politics.¹² Most simply, any change in vote share or in an election outcome will benefit one group and harm another group.¹³ In contrast, most studies on public health, education, or crime intervention typically provide a benefit to some (usually a treatment group), and at least a standard of care to a control group.¹⁴

The second issue is field experiments' combining deception with a lack of informed consent. For many political scientists, deception only refers to deliberately misleading or lying to subjects, and consent only refers to subjects' agreement to be part of a study. Yet in the broader ethics and IRB literatures the two are more closely linked. If some elements of an experiment are hidden from the subjects and not revealed in the consent process, then this is also a form of deception. By this measure, nearly all field experiments are deceptive because the details of the study are usually not disclosed to subjects.

A new ethical challenge of field experiments is that more than deceiving volunteers, subjects are not given a choice to participate or not and often never know that they are participating in a research study. This lack of voluntary and informed participation violates central norms of research ethics, including the Belmont Report's Respect for Subjects¹⁵, the Declaration of Helsinki, and the Nuremberg Code. Yet, such designs are allowed by the Common Rule under limited conditions¹⁶. There is thus an unresolved tension between the norms of voluntary participation and the Common Rule's provisions for studies without consent.

Field experiments raise challenging new ethical questions. Is affecting a political process a form of harm? Is it ethical for researchers to place subjects into minimal risk studies without any informed consent? The issues raised are deep and difficult and unlikely to yield quickly to argument.

An alternative path forward, which I adopt herein, is that of empirical ethics: asking subjects for their judgments on our research.¹⁷ Some have criticized empirical ethics as limited by Hume's is/ought problem: the granting of some moral or "ought" quality to a logical or "is" finding. Even so, scholars note that empirical ethics can

contextualize debates, promote better ethical evaluations, identify unexpected harms, and can shed light on the empirical foundations of ethical questions.¹⁸ For political science, empirical evidence on the opinions of subjects can help resolve the broad challenges posed by field experiments in several ways.

First, empirical ethics can directly contribute to the question of research without informed consent, and its apparent conflict with research ethics norms. Subjects in these studies do not consent, are not debriefed, and typically never know that they were subjects. As a result, we know almost nothing about their feelings about participation in research. Empirical evidence on subjects' views of these designs can fill a critical gap in our understandings of whether forcing subjects into our studies is appropriate or not. If it turns out that subjects widely support such studies and are happy to be included without their explicit consent, then this tension between norms of informed consent and field experiments is eased significantly. Participation is still neither informed nor voluntary, but at least enjoys a "counterfactual consent"—participants would have consented had they been asked. On the other hand, if subjects do not wish to be subjects but are placed into clandestine studies by researchers, then subject participation is involuntary and even fraudulent, and the tension between research ethics and field experiments is heightened considerably.

Second, the opinions of subjects can help inform scholars about subjects' perceptions of harm. A correspondence study that only takes 10 minutes of subject time might seem harmless to a scholar, but a busy potential subject might disagree. In a consenting study, the potential subject could just opt out; in a clandestine study the individual has no choice. Knowledge of what subjects would do if given a choice reveals subjects' perceptions of harm. In addition, they provide insight on whether potential subjects are concerned with social or aggregate harms, or only their individual experience. Lastly, understanding the features of field experiments that are perceived as harmful may help us to find designs that minimize controversy.

Finally, many political science scholars rely on public trust and support: working at public institutions, conducting research with public funds, and using members of the public as subjects and respondents. The trust of these citizen-subjects is critically important. The broader public are our ultimate principals, and research that offends or angers these principals risks harm to the research enterprise. This does not mean that potentially offensive research should never be conducted, but the broader consequences of a reduction in public trust should be part of a cost-benefit assessment.

For all these reasons, I conducted a survey of U.S.-residing adults, asking them to judge a series of hypothetical research designs. I also surveyed scholars. The opinions of

scholars provide a valuable contrast with subjects' opinions, and understanding scholars' collective opinions is a first step toward developing disciplinary norms and guidelines.

The Survey

The survey, conducted in 2015, asked respondents to read short vignettes describing two hypothetical field experiments and to judge their acceptability.¹⁹ One of the vignettes presented an informational field experiment; the other presented a correspondence study field experiment.

Features of the vignettes, described next, were randomized to measure the impact of the critical issues just discussed on the acceptability of hypothetical studies. Vignettes varied deception and participation without voluntary consent, individual and aggregate harm, and research on zero-sum political processes versus research on topics with clearer public benefit.

Informational Field Experiment

In the first vignette, a hypothetical researcher sends flyers to registered voters and then observes their behavior. Several features of the vignette were randomized. The most important was consent: in one version of the vignette, the hypothetical researcher sends flyers to subjects without informing them that they are subjects; in another version, subjects are recruited and consent to participate. The vignette also varied deception—in some cases the flyer was identified as being part of a study, in others it was sent anonymously, and in a third case it was attributed to a non-existent organization. The topic of the study and content of the flyer were alternately presented as reminders to floss, to vote, or that one candidate for elected office had received a DUI. The aggregate impact was varied: the size of the study was reported as either 1,000 or 100,000 subjects, and the study was reported as likely or unlikely to affect an election outcome (only for the turnout and vote-choice versions).

Correspondence Study Field Experiment

The second vignette described a study where the researcher wished to know whether subjects would respond to a request for information. Again, the most important manipulation is consent, this time with three possible treatments. In the first version, the researcher recruits consenting subjects and asks them how they would respond to a hypothetical information request. In a second version of the vignette, the researcher pretends to be a private citizen and subjects never know they are in an experiment. In a third version, there again is no consent, but the subjects are debriefed and offered a chance to have their data deleted from the study. The vignette also varied the topic of the study from a generic investigation of communication to a presumably more valuable investigation of discrimination. The subject population was

randomly described as home sellers, businesses, or elected officials. The aggregate impact and individual burden of the hypothetical study were randomly assigned; size ranged from 500 to 100,000 participants, and the time burden for a hypothetical subject to respond to an informational request was varied from 5 to 60 minutes. Additional details about both vignettes are provided in the online appendix.

Dependent Variables

For both vignettes, subjects and scholars were asked: "To what extent do you agree that it is acceptable to conduct this study?" Responses were coded from 1 (Strongly Disagree) to 7 (Strongly Agree).²⁰

Citizen-subjects were also asked, "Suppose you learned that a study like the one described above had been conducted in your community, and that you were one of the subjects. Which of the following best describes how you would feel about being included in the study?" Subjects could answer, "I would be glad I was in the study", "I would rather not have been in the study", or "I would not care either way". This question was designed to distinguish between subjects' abstract judgements about an experiment and their own feelings as potential subjects. Respondents might judge an experiment as unacceptable, but not care if they were included. Alternatively, they

might think a design acceptable but prefer not to be included in the study.

Sample

The survey of citizen-subjects used 3,000 respondents provided by Survey Sampling International (SSI). The panel was constructed to mirror the U.S. Census. The American Political Science Association generously cooperated with the study, providing a random sample of 14,220 current and former members' e-mail addresses in two waves.²¹ In total, 1,731 of those contacted started the survey, and almost 1,600 completed the four "Agree Acceptable" questions, a response rate of 11%.

Table 1 compares the profile of subjects and scholars surveyed. The citizen-subject sample is roughly representative of the U.S. adult population; the sample of scholars is older and less diverse. 67% of scholar-respondents were ladder rank faculty, 19% graduate students, 5% postdocs, and others were 9%. Among scholars, all major fields were well represented in the survey, and nearly half of scholars had conducted an experiment.²²

Results

Informational Field Experiment

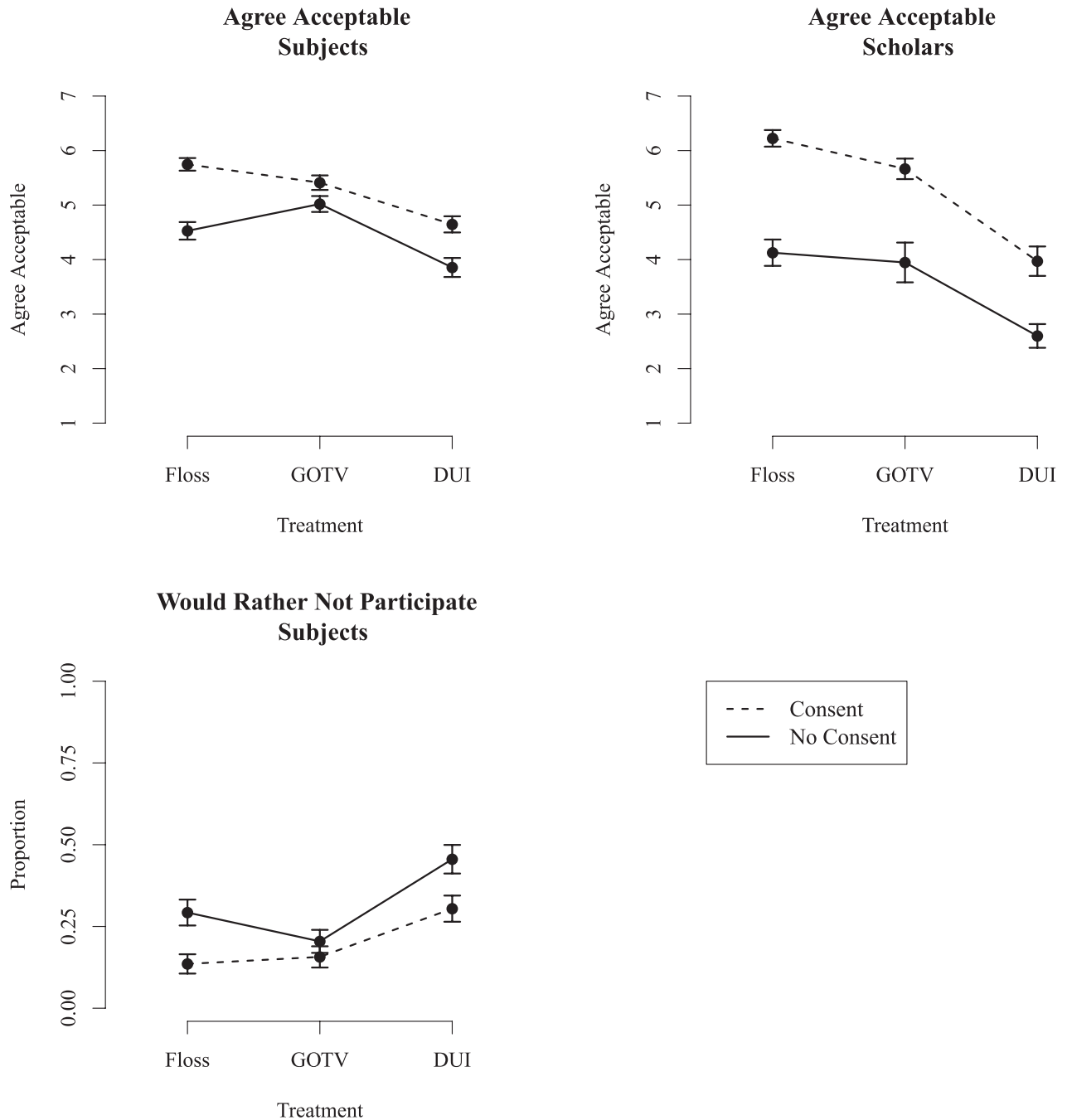
Figure 1 shows the impact of informed consent and research topic on attitudes about informational field

Table 1
Descriptive statistics for all respondents

		Subjects		Scholars	
		%	#	%	#
Gender	Male	48	1443	65	1007
	Female	52	1559	35	540
Age	18-25	12	399	2	43
	26-35	17	576	32	554
	36-45	15	508	25	432
	46-55	16	535	15	263
	56-65	15	508	8	138
	66+	14	476	7	119
	No Response	11	384	11	182
Position	Grad Student	—	—	19	291
	Post-Doc	—	—	5	77
	Assist Prof	—	—	24	376
	Assoc Prof	—	—	21	320
	Full Prof	—	—	22	348
Field(s)	Other	—	—	9	137
	American	—	—	35	614
	Comparative	—	—	38	652
	IR	—	—	20	353
	Theory	—	—	9	157
Ever Experiment?	Methods	—	—	14	245
	Other	—	—	10	174
	No	—	—	48	745
	Yes	—	—	52	808

Note: Additional variables are available in the online appendix.

Figure 1
Attitudes toward informational field experiments



experiments. The left panels show results for subjects; the right panel shows results for scholars. The top panels show results for the “Agree Acceptable” question. In each of these graphs, the X-axis shows the three treatments used in the vignette: flossing, GOTV, or DUI reminders. The Y axis measures agreement that the experiment is

acceptable on a 1–7 scale and the points show mean acceptability with 95% confidence intervals. In the bottom panel, the Y-axis is the percentage of respondents who reported not wanting to participate in a field experiment. For all figures, respondents evaluating an experiment with informed consent are connected with the dashed lines;

respondents that considered the case of a field experiment without consent are connected with the solid lines.

I draw attention to several trends. First, both subjects and scholars are sensitive to the presence or absence of consent. For both groups, and for all treatments, acceptability is significantly lower for the field experiments without consent than for designs with consent. For scholars, mean acceptability (across all three treatments on the x-axis) is 5.33 for an experiment with consenting subjects, but falls to 3.48 for experiments that lack informed consent. Respective figures for subjects are 5.27 and 4.47.

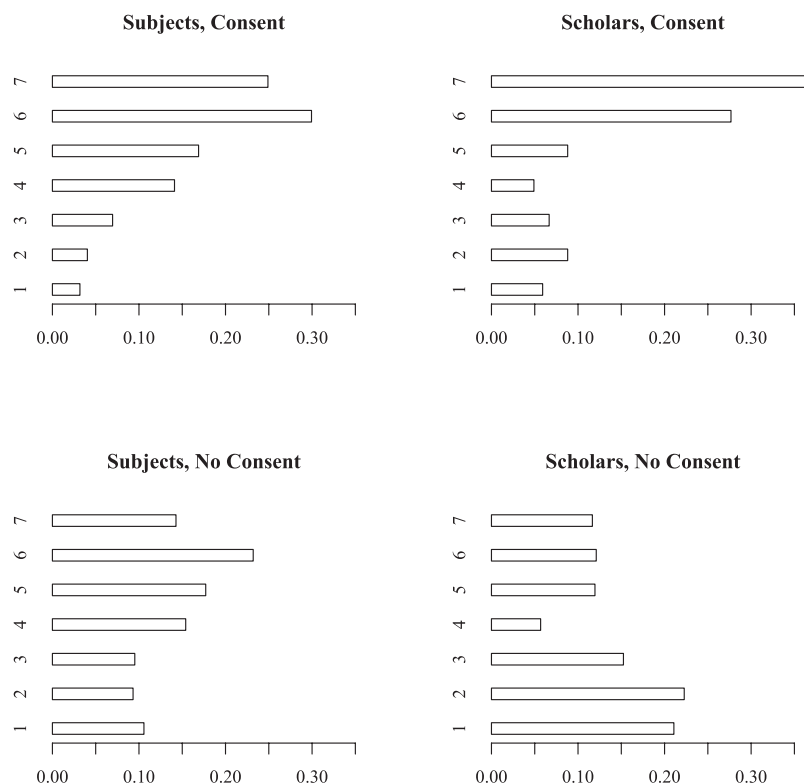
Second, both subjects and scholars are sensitive to the normative value or ambiguity of the topic. I expected highest acceptability for the study with an unambiguous public benefit (flossing), followed by the GOTV and the DUI treatments. For scholars, the expected trend is observed. For subjects, the GOTV treatment is the most acceptable, followed by the flossing treatment, and then the DUI treatment.

Third, although most of the trends are the same for scholars and subjects, scholars are much more sensitive than subjects to the type of study and the presence of

informed consent. For scholars, the mean difference in acceptability between designs with and without informed consent is 1.86; for subjects, the difference is .80. Looking just at studies without informed consent, for scholars, agreement is 1.57 higher for flossing reminders than for DUI reminders; for subjects, the difference is .67.

A look at the underlying response patterns is helpful here, shown in figure 2. These barplots show the distribution of responses to the “Agree Acceptable” question for the informational field experiment. For scholars, the contrast between designs with and without consent is stark. The modal response when considering a study with consent was “7”—strong agreement that the design is acceptable—and fully 74% of respondents are somewhere in the acceptable range (5–7). When the design lacks consent, the most common responses are “2” and “1”—disagreement that the design is acceptable, and the distribution is bimodal—showing division among scholars regarding acceptability of field experiments. In both cases, scholars have opinions on these issues: only about 5% of respondents choose the “neither agree nor disagree” response.

Figure 2
Acceptability of informational field experiments



The distribution of subject responses shows a similar trend, but is much less responsive to the presence or lack of informed consent. For designs with consent, 72% agree the design is at least “Somewhat Acceptable.” Without consent, this figure falls to 55%, still a majority of respondents. The modal response for subjects is “6” (“Agree Acceptable”), both for designs with and designs without informed consent. In the version with consent, only 14% are in one of the “Disagree Acceptable” categories (1–3); this rises to 29% in the vignette where there is no informed consent.

In multivariate models, these same results persist, and the impact of other design features are explored (refer to the online appendix). Both subjects and scholars react negatively to explicit deception—sending a flyer that is attributed to a fake organization significantly lowers the mean acceptability (−.382 for subjects, −.299 for scholars). Scholars are concerned about affecting elections—running an experiment that could affect an electoral outcome reduces acceptability (−.577). For subjects, the risk of affecting an election also reduced acceptability, but the estimated coefficient was smaller and not statistically significant. The size of the hypothetical experiment did not significantly affect acceptability for scholars or for subjects. The interactive models with controls mirrors the original figure: both subjects and scholars respond more to the type of treatment in the presence of consent. This last finding is the opposite of what I expected; in my pre-analysis plan I hypothesized that the type of study would only matter in the absence of consent. In other words, I expected that all designs with informed consent would be highly acceptable, but designs without informed consent would depend on the nature of the study.

Considering control variables for subjects, more educated respondents are significantly more likely to find designs acceptable. For scholars, Americanists found designs more acceptable and Theorists found them less acceptable than did the excluded category (IR). A dummy variable, “Ever Experiment”, was also significant, indicating that experimentalists are generally more accepting of these designs than non-experimentalists. For both subjects and scholars, older respondents and female respondents were less likely to find designs acceptable.

The lower-left graph in figure 1 shows the proportion of respondents who did not want to participate in such an experiment. For the cases with consent, few respondents wish to avoid the GOTV or flossing treatments—just 16% and 14%, respectively, reported that they would rather not participate. For the DUI case, rejection rose considerably, with 30% reporting wanting to avoid the study. Designs without consent had a much higher rejection rate. For the flossing study, the rejection rate was 29% for the version without any consent. The GOTV study saw rejection increase slightly, to 20%. And almost half (46%) would rather not have been in the DUI experiment conducted

without consent. Logistic regressions on an indicator variable for a preference not to participate are in the online appendix, with similar results.

Correspondence Study Field Experiment

Figure 3 shows results for the Correspondence Study Field Experiments, using the same graph format as in the previous example, except that in these figures, the x-axis is the target of the study—hypothetical businesses, politicians, or home owners. In addition, in this study, there were two versions of experiments without consent. In one, the subjects never know they are in a study. In the second, the subjects do not consent to participate, but after the experiment, they are debriefed and given a chance to exclude their data.

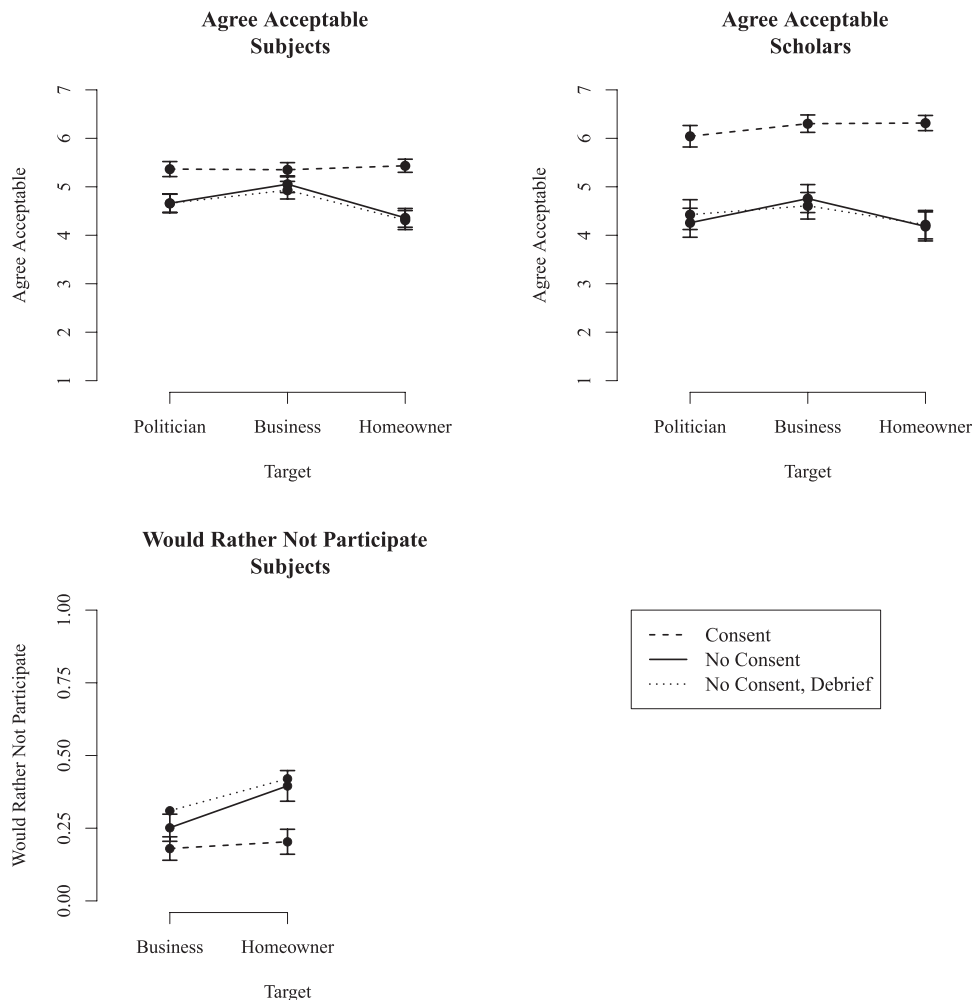
The primary result here is again that designs without consent are less acceptable than those with consent, with a large difference between the two for scholars, and a smaller difference for subjects. For scholars, versions of the design where subjects are fully informed and consenting have uniformly high “Agree Acceptable” scores, with a mean above “6” on the 1–7 scale. For versions with deception and no informed consent (combining versions with and without debriefing), mean agreement falls by 1.82. Subject responses echo those of scholars, but with smaller differences between designs with and without consent (a mean difference of .72).

A second finding is that debriefing has no impact on acceptability or potential participation. For both subjects and scholars, reactions to the design without consent but with debriefing were virtually identical to reactions to the design without consent and without any debriefing. The dotted and solid lines track almost perfectly, and are never statistically distinguishable. Debriefing has been proposed as a form of “Deferred Consent”²³ and it is required “whenever appropriate” by the Common Rule, but does not increase acceptability or willingness to participate.

A third finding is that there is only a modest impact of the hypothetical target subject population on acceptability. Although studies of public officials are exempt under the current version of the Common Rule, treating them is actually less acceptable than treating business owners, for both scholars and subjects. As expected, home sellers are the least acceptable hypothetical target for such studies, though the difference between populations is modest for both samples. The graph suggests an interaction: skipping consent appears less acceptable when targeting private homeowners than when targeting politicians or business owners.

In multivariate models (available in the online appendix), all these trends persist and the effects of several other variables are tested. The normative value of the topic is relevant—designs that study discrimination are significantly more acceptable than those that study

Figure 3
Attitudes toward correspondence study field experiments



communication, customer service, or constituency service (estimated coefficient on “Discrimination Topic” was roughly +.3 for both subjects and scholars).²⁴ A higher burden on subjects reduces acceptability for both groups. The size of the study and debriefing did not affect acceptability for subjects or for scholars.

Finally, the lower-left panel in figure 3 shows the proportion of subjects preferring not to participate in such studies, by target and deception. In this case, the follow-up question was only asked for the homeowner and business versions of the vignette. As with the informational field experiment, rejection is low in the case of informed consent and varies little across target. For the business version of the design, 18% reported preferring not to be in the study. For the homeowner version, that rose slightly to 20%. However, for the version without informed consent, where the researcher pretends to be a potential customer or

potential home-buyer, rejection is much higher, at 28% and 41%, respectively.²⁵ The logistic regressions with and without controls (available in the online appendix) largely reiterate these findings.

Results here echo findings from the last section. Consent has a significant effect on subject and scholar attitudes. The normative value of the study affects both groups’ attitudes—studies of discrimination are more acceptable than those of communication. Most importantly, large numbers of subjects would rather not be included in some studies without consent.

Discussion

I offer four primary empirical findings. First, both subjects and scholars react negatively to experiments without consent and to all forms of deception. For both populations, removing consent or adding deception

significantly reduced mean acceptability scores, even for minimal risk and minimally intrusive experiments. For scholars, designs without informed consent were polarizing and reveal divisions among political scientists.

Second, the nature of the research affects judgements. Scholars and subjects were more tolerant of research with clear public benefit than of research with more ambiguous benefit. Respondents' comments on the discrimination version of the correspondence study expressed an interest in seeing results and an appreciation for the importance of the research. Comments on the vote choice experiment included expressions of suspicion that the study might be an attempt to manipulate the electoral process.

Third, subjects appear less concerned about these issues than scholars. Subjects were only modestly responsive to treatments and on average lukewarm toward many of the experimental designs. In contrast, scholars reacted strongly to small design changes. As a result, subjects' opinions moved in a narrow band, while scholars' opinions often jumped sharply above and below subjects in response to design changes. This may indicate that subjects do not care as much about these issues, that subjects paid less attention to the survey, or that they have not thought about these issues as much as scholars have.

Lastly, the most important takeaway is that, for some designs, many subjects would rather not be subjects. Opposition to participation was low in cases where there was no deception and where the topic had clear normative value. In studies without any consent and on more ambiguous topics, this increased to nearly half of the respondents.

These empirical findings should prompt some sober reflection by political scientists. Many of our subjects are placed into our studies against their will. In some designs, most respondents were willing to participate in research as long as they were consenting, and it was the lack of consent that prompted their rejection. In other designs, subjects did not like the study, did not want to participate consenting or not, and the prospect of being forced into a study only increased rejection.

How should the field proceed? One response is to defend the status quo, pointing to the quality of the science and the fact that most political science field experiments have IRB approval. But IRB approval is neither ethical approval nor legal absolution,²⁶ and one may question the scientific advantages of field experiments.²⁷ More importantly, if we justify forcing individuals to be subjects against their will, based on the benefits to our research, we may join the ranks of the most infamous of medical research disasters.

I'll suggest three practical ways to move forward. First, we might find a middle ground by seeking creative forms of consent. Humphreys²⁸ proposes several alternative forms, including proxy consent, superset consent, and

several others. Bioethicists have proposed using citizen panels to evaluate research when the issues are too complicated for a simple informed consent script.²⁹ Medical research on emergency medicine—where subjects are often unconscious and unable to consent—has used community information campaigns and given individuals a chance to opt out of research, should they wind up unconscious in an emergency room. In political science, Zimmerman³⁰ deployed a similar model in Africa, informing the community about the research through media outlets. Another possibility is recruiting long-term panels of subjects who agree to participate in clandestine IFE or CSFE, without telling them all the details of the research or when treatments might occur.

Second, we can minimize harm to subjects, society, and research by following some best practices suggested by this research. Above all, the results support striving to use informed consent whenever possible. We should exhaust learning from experiments with consent before using designs without consent. When scholars decide to proceed without consent, they should defend the design in terms of benefits versus harms, recognizing the real risks to subjects, society, and to the research enterprise. In addition, we should design field experiments to minimize subject rejection and harm, following several principles.³¹

Do good. Respondents had higher tolerance for field experiments on topics perceived as being clearly in the public interest. Researchers conducting interventions without subjects' consent should focus on areas where the treatment and outcome offer clear public benefit. In addition, scholars need to do a better job of explaining the value of basic research, and address suspicion that our studies are an attempt to manipulate election outcomes. There is an ethical need to conduct basic research, even if the knowledge can be used for ill.³² Such arguments can be extended to topics of voter persuasion and negative campaigning, and may help justify such research to otherwise suspicious subjects.

Tread lightly. Minimize impacts on political processes and subjects. In some IFEs political scientists have out-campaigned the real politicians—outspending real candidates and contacting more voters than did the real candidates. Treading lightly implies conducting a power analysis and minimizing the size and burden of the study.

Confess and compensate. Debrief subjects. Debriefing shows respect for subjects, provides useful data on subjects' opinions, allows scholars the opportunity to explain and defend the research to subjects, and makes scholars accountable for their research. Finally, more than debriefing, compensate subjects post-study. This shows respect for subjects' time, may assuage opposition to our studies, and provides a financial constraint on scientists' enthusiasm for massive interventions.

Lastly, we need more empirical research on the ethics of our work. My study has many limitations and leaves

many questions unanswered. Opposition to designs might disappear if scholars could explain the aims and importance of the research, and the reasons for the chosen approaches. Or subjects might welcome clandestine field experiments if they received post-study compensation. These results might not hold in other countries or contexts, or even with other question wordings. Finally, there are many issues not addressed herein, including field experiments examining illegal activity, conducted in authoritarian regimes, or developed with third-party organizations. For all these reasons, this study should not be seen as the last word, but merely as some introductory remarks in an overdue conversation in which all should participate.

Notes

- 1 Morton and Williams 2010; Bositis and Steinel 1987; Druckman et al. 2006; McDermott 2002; Desposato 2016a.
- 2 However, as King and Sands 2015 point out, IRB approval may not protect scholars from the consequences of their interventions.
- 3 Michelson 2014.
- 4 One reason we do not know is because subjects in political science field experiments are rarely debriefed.
- 5 Political scientists are involved in field experiments that range beyond typical political science topics. In addition, scholars are partnering with non-academic parties who are implementing the experiments. Such situations raise issues beyond the scope of this manuscript. See Nickerson and Hyde 2016; Humphreys 2011; Tucker 2011; Hyde 2011; De La O 2011.
- 6 There are hundreds of published IFEs and CSFEs. See Findley, Nielson and Sharman 2014 ; De La O 2011; Humphreys 2011; Grose 2014; John 2017; Green and Gerber 2015 for some examples of field experiments and a discussion of their benefits.
- 7 Of course, there are countries and contexts where an IFE could be inappropriate, illegal, or dangerous to researcher and subject, considerations discussed in Driscoll 2016 and Michelson 2014.
- 8 Mai-Duc 2017; Theriault Boots 2014.
- 9 Gelman 2010.
- 10 The case discussed in Kifner 2001 led to a lawsuit against Columbia University.
- 11 McDermott 2017.
- 12 Zimmerman 2016, ch. 12; Humphreys 2014.
- 13 Changing vote share without affecting an election might not be viewed as harmful, but such changes have other downstream consequences for candidates and parties, affecting fundraising, nominations, and candidate recruitment.
- 14 This is not an argument that political processes should not be studied. Indeed, as Hatemi and McDermott 2011 have pointed out, it would be unethical not to study normatively difficult topics.
- 15 Teele 2014.
- 16 U.S. Department of Health and Human Services 2015.
- 17 Blomquist 1975.
- 18 Hume 1896; Borry, Schotsmans, and Dierickx 2008; Hope 1999; De Vries and Gordijn 2009.
- 19 The survey in fact had four vignettes. Two were field experiments, described in this article. Space limitations prevent me from discussing the other two. One omitted vignette examined attitudes regarding international research conducted without permission from the host government. The other examined attitudes about the use of deception in laboratory experiments.
- 20 This question follows the model of previous research, for example, Ludman et al. 2010. Technically, the question suffers from a lack of balance and from potential acquiescence bias; Schuman and Presser 1996. Both should bias acceptability upward, suggesting that actual acceptability should be lower for all vignettes.
- 21 In both waves, I excluded colleagues who had tested the survey. There is a small risk that some APSA subjects may have responded twice to the survey. Details are in the online appendix.
- 22 The high percentage of scholar-respondents that have conducted an experiment suggests a disproportionately high response rate by experimentalists compared to non-experimentalists. Given that experimentalists were generally more favorable to all designs and that experimentalists were overrepresented in the sample, it may be that overall approval of experiments among political scientists is in fact lower than reported here.
- 23 Humphreys 2015, 13.
- 24 For subjects, this result persisted in interactive models. Their optional comments also expressed more interest and support for studies on discrimination than for studies on communication. For scholars, the topic did not matter in studies conducted with informed consent, but "Discrimination" had an even larger impact on acceptability for studies conducted without consent."
- 25 These are overall percentages combining both consent, and consent with debrief.
- 26 King and Sands 2015.
- 27 Some scholars disagree with the degree of concern with the Hawthorne effect. Others have found nearly identical results in corruption studies conducted in laboratories and in the field; Armantier and Boly 2008. Field experiments are naturally embedded in culture and context and may be far less generalizable than laboratory experiments.
- 28 Humphreys 2015.

- 29 Koenig 2014.
 30 Zimmerman 2016.
 31 I discuss these more fully in Desposato 2016b.
 32 Hatemi and McDermott 2011.

Supplementary Materials

To view supplementary material for this article, please visit <https://doi.org/10.1017/S1537592717004297>

References

- Armantier, Olivier and Amadou Boly. 2008. "Can Corruption Be Studied in the Lab? Comparing a Field and a Lab Experiment." *CIRANO Scientific Series* 2008s-26. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1324120.
- Blomquist, Clarence. 1975. "The Teaching of Medical Ethics in Sweden." *Journal of Medical Ethics* 1: 96–8.
- Borry, Pascal, Paul Schotsmans, and Kris Dierickx. 2008. "The Origin and Emergence of Empirical Ethics." In *Empirical Ethics in Psychiatry*, ed. Guy Widdershoven, John McMillan, Tony Hope and Lieke van der Scheer. Oxford: Oxford University Press.
- Bositis, David A. and Douglas Steinel. 1987. "A Synoptic History and Typology of Experimental Research in Political Science." *Political Behavior* 9(3): 263–84.
- De La O, Ana L. 2011. "The Experimental Turn in the Study of Democratization." *APSA Comparative Democratization* 9(3).
- De Vries, Rob and Bert Gordijn. 2009. "Empirical Ethics and its Alleged Meta-Ethical Fallacies." *Bioethics* 23(4): 193–201.
- Declaration of Helsinki. 2008 [1964]. Available at <http://www.wma.net/en/30publications/10policies/b3/17c.pdf>.
- Desposato, Scott. 2016a. "Introduction." In *Ethics and Experiments: Problems and Solutions for Social Scientists and Policy Professionals*, ed. Scott Desposato. Routledge Series in Experimental Political Science. New York: Routledge.
- . ed. 2016b. "Conclusion and Recommendations." *Ethics and Experiments: Problems and Solutions for Social Scientists and Policy Professionals*. Routledge Series in Experimental Political Science. New York: Routledge.
- Driscoll, Jesse. 2016. "Prison States and Games of Chicken." In *Ethics and Experiments: Problems and Solutions for Social Scientists and Policy Professionals*, ed. Scott Desposato. Routledge Series in Experimental Political Science. New York: Routledge.
- Druckman, James N., Donald R. Green, James H. Kuklinski, and Arthur Lupia. 2006. "The Growth and Development of Experimental Research in Political Science." *American Political Science Review* 100(4): 627–35.
- Findley, Michael G., Daniel L. Nielson, and J. C. Sharman. 2014. *Global Shell Games: Experiments in Transnational Relations, Crime, and Terrorism*. New York: Cambridge University Press.
- Gelman, Andrew. 2010. "\$63,000 Worth of Abusive Research ... or Just a Really Stupid Waste of Time?" Available at <http://andrewgelman.com/2010/05/06/63000worthof/>
- Green, Donald P. and Alan S. Gerber. 2015. *Get Out the Vote: How to Increase Voter Turnout*. 3rd ed. Washington, DC: Brookings Institution Press.
- Grose, Christian R. 2014. "Field Experimental Work on Political Institutions." *Annual Review of Political Science* 17: 355–70.
- Hatemi, Peter and Rose McDermott. 2011. "The Normative Implications of Biological Research." *PS: Political Science and Politics* 44(2): 325–329.
- Hope, Tony. 1999. "Empirical Medical Ethics." *Journal of Medical Ethics* 25: 219–20.
- Hume, David. 1896 [1739]. *A Treatise of Human Nature*. Available at <http://oll.libertyfund.org/titles/hume-a-treatise-of-human-nature>.
- Humphreys, Macartan. 2011. "Ethical Challenges of Embedded Experimentation." *APSA Comparative Democratization* 9(3).
- . 2014. "How to make field experiments more ethical." *The Monkey Cage*, November 2.
- . 2015. "Reflections on the Ethics of Social Experimentation." *Journal of Globalization and Development* 6(1): 87–112.
- Hyde, Susan D. 2011. "Anybody's Luck? Natural Experiments in Democratization." *APSA Comparative Democratization* 9(3).
- John, Peter. 2017. "Field Experiments on Political Behavior." *Politics: Oxford Research Encyclopedias*. DOI:10.1093/acrefore/9780190228637.013.230.
- Kifner, John. 2001. "Scholar Sets Off Gastronomic False Alarm." *New York Times*. Available at <http://www.nytimes.com/2001/09/08/nyregion/scholar-sets-off-gastronomic-false-alarm.html>; accessed April 7, 2015.
- King, Gary and Melissa Sands. 2015. "How Human Subjects Research Rules Mislead You and Your University, and What to Do About it." Working Paper, Harvard University.
- Koenig, Barbara. 2014. "Have We Asked Too Much of Consent?" *Hastings Center Report* 44(4): 33–34.
- Ludman, Evette J., Stephanie M. Fullerton, Leslie Spangler, and Susan Brown. 2010. "Glad You Asked: Participants' Opinions of Re-Consent for dbGap Data Submission." *Journal of Empirical Research on Human Research Ethics: An International Journal* 5(3): 9–16.
- Mai-Duc, Christine. 2017. "A Letter Sent to Some L.A. Voters Sought to Shame Them for Their Voting Records—And No One Knows Who Sent It." *Los Angeles Times*. Available at <http://www.latimes.com/politics/la-pol-ca-voter-shaming-mailer-20170516-htmlstory.html>.
- McDermott, Rose. 2002. "Experimental Methods in Political Science." *Annual Review of Political Science* 5: 31–61.

- . 2017. "Ethics in Field Experimentation: A Call for Best Ethical Practices and a New Standard-Respect for Society." Working Paper, Brown University.
- Michelson, Melissa R. 2014. "Messing with Montana: Get-out-the-Vote Experiment Raises Ethics Questions." *The New West*, October 25. Available at <https://thewpsa.wordpress.com/2014/10/25/messing-with-montana-get-out-the-vote-experiment-raises-ethics-questions/>.
- Morton, Rebecca B. and Kenneth C. Williams. 2010. *Experimental Political Science and the Study of Causality: From Nature to the Lab*. New York: Cambridge University Press.
- Nickerson, David W. and Susan D. Hyde. 2016. "Conducting Research with NGOs: Relevant Counterfactuals from the Perspective of Subjects." In *Ethics and Experiments: Problems and Solutions for Social Scientists and Policy Professionals*, ed. Scott Desposato. Routledge Series in Experimental Political Science. New York: Routledge.
- Schuman, Howard and Stanley Presser. 1996. *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context*. Thousand Oaks: Sage Publications.
- Teele, Dawn Langan. 2014. "Reflections on the Ethics of Field Experiments." *Field Experiments and Their Critics: Essays on the Uses and Abuses of Experimentation in the Social Sciences*. New Haven, CT: Yale University Press.
- The Nuremberg Code. 1947.
- Therhault Boots, Michelle. 2014. "Alaska voters upset about public shaming mailers, but experts say they work.. *Anchorage Daily News*, September 28. Available at <https://www.adn.com/politics/article/mailings-use-public-shaming-tool-motivate-voters-anger-experts-say-they-work/2014/10/28/>.
- Tucker, Joshua. 2011. "Survey Experiments: What They Are, What They Can Do, and Why They Are Especially Important in New Democracies." *APSA Comparative Democratization* 9(3).
- U.S. Department of Health and Human Services. 2015. "Federal Policy for the Protection of Human Subjects." Available at <https://www.hhs.gov/ohrp/regulations-and-policy/regulations/common-rule/index.html>.
- Zimmerman, Brigitte. 2016. "Information and Power: Ethical Considerations of Political Information Experiments." In *Ethics and Experiments: Problems and Solutions for Social Scientists and Policy Professionals*, ed. Scott Desposato. Routledge Series in Experimental Political Science. New York: