Subjects' and Scholars' Views on Experimental Political Science

Scott Desposato*

September 22, 2016

*Thanks to Seth Hill, Seonghui Lee, Steve Smith, Marco Steenberg, Dawn Teele, Liz Zechmeister, the Institut für Politikwissenschaft, Universität Zürich, and the American Political Science Association. This research was approved by the UC San Diego Human Research Protections Program. Analysis and design were pre-registered at Open Science.

Abstract

Recent controversies involving political science reveal disagreement regarding the ethics of experimental social science. Two central issues are the use or lack of informed consent in field experiments and the appropriate review of international experiments. In this paper I present results from a survey of scholars and of the public on the ethics of typical political science experimental designs. In the survey, the research designs were randomly varied to identify the most controversial features of political science experiments. Both scholars and subjects react negatively to deception and to field experiments without informed consent, especially when the research project is normatively ambiguous. In some cases, almost half of subject respondents reported that they would rather not be in a typical field experiment without their consent. Further, both types of respondents oppose conducting experiments overseas without any local approval.

1 Introduction

The dramatic growth in experimental political science has led to new and unexpected ethical controversies. Some of these mirror existing debates in bioethics, but others are new to research ethics paradigms. There is apparent disagreement among political scientists regarding how the field should proceed. Many of the new issues do not have obvious solutions, yet may still risk harm to subjects, bystanders, researchers, and the profession.

My goal in this paper is not to debate the theoretical ethics of experimental design, but to examine "empirical ethics" - investigating how subjects and scholars perceive experimental research, identifying the most problematic elements of current research trends, and finding designs that minimize risk and allow research to continue.

To that end, I report herein on a survey on experimental ethics administered to citizen-subjects and to scholars. Participants read several vignettes that described typical political science research projects, and then reported their judgements on the designs. Key features of each hypothetical research project were randomly varied to allow identification of the most controversial and contentious aspects of experimental political science. Two consistent results for both citizen-subjects and scholars are 1) disagreement over the acceptability of experiments that lack informed consent, and 2) rejection of international experiments conducted without the approval of local authorities. In addition, both subjects and scholars are sensitive to the normative value of the research; greater risk is tolerated for some research topics than others. Lastly, scholars' judgements about hypothetical experiments are generally much more critical and sensitive to features of the research design than are subjects' judgements. Some of the findings from this survey will only apply to experimental research, but others may contribute to observational and qualitative research.

The paper proceeds in three steps. In the next section, I discuss the rise of experimental political science, the new ethical challenges, and the potential contributions of empirical ethics. In Section 3, I present the survey experiment and discuss hypotheses. In Section 4, I present results from the survey. Section 5 concludes.

2 New Ethical Challenges

One of the most dramatic changes in political science over the last 25 years is the experimental revolution. Twenty-five years ago, political science experiments were rare, restricted to a handful of subfields, and typically involved undergraduates or university staff playing economic games or watching videos in empty classrooms. Today scholars in all subfields are using experimental methods to test theories and measure causal effects. Further, more than just games in US university classrooms, scholars are conducting experiments in almost every country and are increasingly using large-scale field experiments, sometimes with thousands of uninformed and unconsenting subjects.[1] These changes in the discipline are not without their critics(Teele, 2014), but experiments have clearly opened a new research frontier and are a powerful set of tools for answering some questions.[2]

This experimental revolution has been accompanied by an unexpected new set of ethical controversies, many of which do not fit neatly into existing human subjects paradigms. I focus herein on the two that are most pressing for the discipline: field experiments without informed consent and international experiments without local authorities' approval. I'll provide an overview of the disagreements involving each in the following paragraphs.[3]

2.1 Field Experiments

Field experiments are manipulations conducted "in the real world" rather than in a laboratory. These experiments often involve treatments administered without the informed consent of subjects. For inference and theory testing, these designs have obvious advantages. They provide estimates of treatment effects in practice, rather than just in a laboratory setting.

Political scientists are involved in diverse field experiments involving development, international regulation, health, and education, as well as more typically political science studies

[1] See Morton and Williams (2010); Bositis and Steinel (1987); Druckman et al. (2006); McDermott (2002); Desposato (2016*b*) on the rise of experimental methods.

[2] It is important to note that neither randomization nor experimentation are required for many ethical issues. Many observational and qualitative studies are confronting similar and some very different ethical challenges.

[3] For a more complete discussion of the issues associated with the experimental revolution, see Desposato (2016*b*).

of turnout, vote choice, or elected official responsiveness(Olken, 2010; O, 2013; Findley, Nielson and Sharman, 2014; Lieberman, Posner and Tsai, 2014; Green, 2004; Loewen and Rubenson, 2011; Mendez and Grose, 2014; Broockman, 2013). In some of these, the political scientist is the principal investigator, and designs, funds, and controls implementation of the experiment. In others, the scholar is merely an advisor to a campaign, a government, or an aid agency seeking to improve service delivery. [4]

I focus 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] These two designs are being widely used in political science and illustrate some of the most important ethical challenges facing the discipline. In particular, they raise questions about aggregate versus disaggregated risk, the normative ambiguity of many political science studies, and to varying degrees, deception and informed consent. I'll briefly discuss each type of field experiment and the nature of disagreement over their ethical status.

Informational Field Experiments - IFE's

Informational field experiments involve assigning subjects to receive information, then observing behavior. In political science, these are usually election studies. Scholars might send information about political candidates or voting procedures to subjects, then observe whether they vote, or for whom they vote. In cases where behavior of individuals cannot be directly observed - voting with secret ballots, or turnout in countries without public turnout records - scholars treat precincts instead of individuals, with all households in the same precinct receiving the same message. A key feature of these designs is that subjects often never know that they are subjects - the designs do not have any consent, and often employ deception as mailers may purport to be from a non-existent group or contain other mislead-

[4]Nickerson and Hyde (2016) discuss the ethical issues involved in these types of studies, called "third-party interventions".

[5]Audit studies send real trained confederates to interact with unknowning subjects. Examples include interviewing for rental housing or employment. Correspondence studies involve submitting applications or other documents on behalf of fictitious individuals.

ing information. This surreptitious design allows scholars to avoid to measure exactly how different types of information affect behavior - actual turnout, versus turnout intentions.[6]

For supporters of these studies, the defense seems easy and obvious. IFE's without informed consent are legal allowed under the Common Rule's (45 CFR Part 46, 116.d).[7] IFE treatments are usually typical of other campaign messages subjects receive, and are equally low-risk - the biggest threat of receiving a mailer might be that of a paper cut. Concerns about changing election outcomes are probably overstated. If it were so easy to affect an election, then candidates and consultants would already be doing so! Lastly, the treatments can be defended as normatively valuable: providing factual information about candidates to voters increases citizen knowledge and accountability; encouraging turnout and participation might strengthen democracy.

However, from a different perspective such designs could be problematic. The risk of affecting an election outcome could be significant, especially when studying contexts where candidates have fewer resources - perhaps outside the United States, or in a local election. In such contexts, the researcher might in fact run the biggest campaign. A graduate student with $10,000 in NSF funding might have a larger budget than all candidates in some contests.

[6]Some comment is required on the distinction between deception and consent. In conversations with political scientists, there seems to be a consensus that deception only refers to deliberately misleading or lying to subjects, and that consent only refers to subjects' agreement to be part of a study. Thus one could imagine deception being used in studies with or without consent.

In the broader bioethics and IRB literatures, however, the two are more closely linked. If some elements of an experiment are hidden from the subjects and not revealed in the informed consent, then this is deception by omission (failure to disclose some relevant feature of the study) versus deception by commission (deliberate provision of false information). By this measure, nearly all field experiments are deceptive because in most cases, the details of the study are not disclosed to subjects. Indeed, the survey experiment presented in this paper would be deceptive because the consent form did not explain that the scenarios were randomized.

In this paper I will use a narrow definition of deception, referring to the provision of explicitly false information to subjects as deception.

[7]This section allows researchers to skip informed consent when four conditions are met: 1) the research is no more than minimal risk 2) skipping informed consent will not affect the rights and welfare of the subjects 3) the research could not practicably be carried out without the waiver 4) subjects will receive additional information after participation. Most political science field experiments of which I am aware do not appear to have ever debriefed subjects in compliance with item (4).

The researchers in the Montana study had a reported $350,000 for their study, which was nearly as much as the combined fundraising of all four candidates in the election they studied. One of the candidates they targeted had raised only $6,100.(Cohen, 2014; National Institute on Money in State Politics, 2014)

Further, even if a study does not change an election outcome, interventions in elections might have other affects on candidates' ability to raise funding, candidate willingness to run again, and candidate recruitment. And while providing factual information to voters is valuable, exactly what constitutes factual and unbiased information is not clear. For example, a negative advertisement study might inform voters that a candidate was accused of corruption, but fail to mention that the allegations were dismissed. The treatment might be factual but misleading.

Correspondence Study Field Experiments - CSFE's

A different type of field experiment involves additional deception and contact with subjects. With a correspondence study, more than just sending information to a subject, the scholar (or her assistants) interact with subjects. Typically, scholars contact politicians or government officials, pretending to be constituents with a policy concern or a problem with a government program. In other studies, scholars send fake resumes to businesses, randomizing features of fake applicants to learn about the characteristics that affect employment. Closer to home, scholars might also pretend to be prospective or current graduate students and contact faculty requesting information about their university or asking for replication datasets. Such designs are especially useful for revealing socially unacceptable behavior (like discrimination).[8]

Disagreement about the ethics of such designs is similar to questions raised about informational field experiments. Again, the defense of such designs is that the risk or cost to individuals are typically low - a few minutes of subject time - and the requests made of subjects are typical of what subjects would encounter as part of their normal employment.

[8]Correspondence studies are distinguished from audit studies in that the latter are usually defined as involving sending trained confederates to interviews or apply for housing or positions. Correspondence studies interact with subjects via email or letters, and have less risk of any pollution of results due to personalities or other individual characteristics besides the ones that scholars are seeking to manipulate.

As with IFE's these types of studies may be permitted under the Common Rule's waiver of informed consent. In addition, when subjects are public officials, studies are exempt.[9]

However, while the disaggregate cost of such experiments may be low, there are potentially large aggregate social costs. Responding to a constituent request might take only 15 minutes, but when treating 1,000 local officials, schoars are using 15,000 minutes of public labor. And again, these designs violate norms of informed consent and respect for subjects' autonomy, by deceiving them into participating in research. Subjects believe they are acting with some potential benefit - they seek a new employee or more constituent support. In fact, the true expected value of the interaction is negative - there is no payoff, only the cost of their time. And, as with the IFE's, when subjects detect these deceptions, they are often very upset.[10]

Broader Issues for Field Experiments

One way to understand disagreement about both Informational Field Experiments and Role-Playing Field Experiments is that these designs combine three problematic issues: aggregate versus disaggregate harm or risk, the zero-sum nature of elections, and the combination of deception and informed consent.

The first issue is whether we should be concerned with disaggregated, or individual risk, or with aggregate risk. The individual risk of harm in most IFE's and CSFE's is trivially small, but aggregate effects are potentially large. These aggregate risks include third-party harms and spillovers - affecting other citizens in a political system. They also include potential harms to the original subjects where individually they are not harmed by receiving an election flyer, but collectively they may be harmed by a change in election outcome.

These broad spillovers and aggregate effects might not be so important if our research were always beneficial to subjects. But many political science projects are normatively

[9]For DFE's dealing with appointed officials, elected officials, or candidates for elected office, research is also exempt from the requirements of informed consent under 45 CFR Part 46, 101.b.3.i. See Grose (2016); Malesky (2016) for a discussion of ethical issues involved when treating public officials.

[10]For an example involving scholars, see Gelman's (2010) discussion and subsequent comments in response to a deceptive email field experiment.

ambiguous. This is especially the case with interventions in campaigns, where the zero-sum nature of elections makes these types of studies unique when compared with other research. Any impact of an experiment on an election benefits one candidate or constituency and harms another. In contrast, an informational experiment to promote flossing or water conservation doesn't have the same ambiguity; improving outcomes for one household need not harm another.[11]

Lastly, even if legal and low risk, the lack of informed consent and use of widespread deception are controversial, as they violate central norms of human subjects research: the voluntary and informed participation of subjects. The primacy of the informed and voluntary consent is the first point of the Nuremberg Code, is also in the Declaration of Helsinki, and included as part of respect for subjects in the Belmont Report(*The Nuremberg Code*, 1947; *Declaration of Helsinki*, 2008 [1964]; The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979). Indeed, the document that allows us to skip informed consent - the Common Rule - is a set of federal regulations - not an ethics document(U.S. Department of Health and Human Services, 2015). Finally, when subjects realize they are in experiments without their consent, subjects often respond with anger and resentment.[12]

Field experiments without informed consent are often legal - but they may not be ethical. How should political scientists proceed?

2.2 Foreign Review of International Experiments

The second ethical challenge is one unique to experiments conducted outside a scholar's home country. Many countries have rules that govern responsible research and most require some form of local review of research: the host country must review and approve any research projects. Many scholars, however, ignore these rules, flying in on tourist visas and conducting "black-ops" experiments without any input from local human subjects' committees. It is

[11]One can construct winners and losers for this type of research through extreme rhetorical contortions. For example, water conservation might reduce income for water utilities and their shareholders. Increased flossing might reduce dentists' revenue.

[12]In response to a very mild Facebook manipulation of users' feeds, thousands of users criticized the study. Facebook eventually apologized. For other examples, see Desposato (2016*a*).

easy to understand why scholars usually take this approach: US-based universities often do not even ask about local approval when reviewing experiments that will be conducted overseas; our primary funding vehicle, the NSF, historically did not ask about local approval; and obtaining such approval can be very difficult, ethically challenging, and agenda-restricting. Regarding this last point, political science is uniquely positioned to alienate powerful actors - research on corruption, democratization, and similar topics are unlikely to appeal to many governments, who are thus unlikely to approve much political science research. Full compliance with foreign rules would in some countries severely restrict our agendas to topics that authoritarian governments find appealing - perhaps on effective methods of oppressing opposition movements or on efficient censorship(Tran, 2015). Thus, full compliance with some countries' rules might itself be unethical. In other cases, local review processes might be long and arduous, or unreasonably demanding.

Most of the time, research is innocuous and no one is harmed by our skipping local review. But conducting research illegally overseas certainly comes with risk to scholars, their local collaborators, and to all scholars working in this area. In the case where funding is provided by the US' National Science Foundation, it is easy to imagine a diplomatic crisis if a scholar supported by the US government were running illegal experiments. Should we always comply with host countries' rules? If not, when can we ignore them, and what are best practices for "illegal" research? These are not easy questions, and they do not have clear answers. Scholars have proposed substituting informal local review - where a trusted and contextually sensitive colleague reviews the design. Others have suggested fully engaging the review process and even working with local authorities to make it manageable. Others argue that we should just continue to fly under the radar, maintaining an uneasy equilibrium that relies on an inattentive government and informal norms among scholars that keep research from becoming too high-profile.

2.3 Empirical Ethics

Political science is engaging in research that is legal, but may be unethical, as well as other types of research that are illegal, but may be ethical. The questions being raised are deep

and difficult, and unlikely to yield quickly to ongoing dialogue. However, these questions need engagement and resolution - as they pose real risks to subjects and to the research enterprise.

I submit that one way forward involves "empirical ethics" - asking what our subjects think about our research.[13] Certainly public opinion is not equal to ethics, and mass opinion can support unethical behavior. But understanding what subjects think about our designs is important for at least reasons. One is that, if our subjects strongly reject what we are doing to them, then we probably should not be doing it. In particular, if subjects would rather not be in the kinds of field experiments many are conducting, this suggests that field experiments are forcing subjects into studies against their will and violating central norms of respect for subjects.[14] In addition, to the extent that we understand the source of subjects' beliefs, we may be able to assuage rejection of our research by adapting our designs and by educating the public about what we are doing and why we are doing it.

Second, there are practical consequences to ignoring subjects' preferences. One is that we jeopardize public trust in the research enterprise, which may deter participation in all types of studies, not just political science. More instrumentally, many of us are employed at public universities and have research funding from public sources. Our subjects are also our principals - they elect the leader-agents that make laws, set budgets, and regulate research. Disregard and disrespect for our subjects could easily produce a broad backlash against all political science research. Such consequences deserve consideration.

As indicated in the title, besides subjects, I will also examine the opinion of scholars.

[13]Empirical, or descriptive ethics is a subfield of ethics, the goal of which is to assess ethical belief systems and their development(Blomquist, 1975). Empirical evidence can inform theoretical debates, help practitioners adapt their ethics to patient context, and provide insight on clinical thinking. See Borry, Schotsmans and Dierickx (2008) for a summary of the field. Designs such as this one have been used in many other fields. See, for example, Roberts et al. (2002).

[14]From the Belmont Report: "Respect for persons requires that subjects, to the degree that they are capable, be given the opportunity to choose what shall or shall not happen to them. This opportunity is provided when adequate standards for informed consent are satisfied."; the Declaration of Helsinki (2008 [1964]), "...no individual capable of giving informed consent may be enrolled in a research study unless he or she freely agrees."; and the Nuremberg Code (1947), "The voluntary consent of the human subject is absolutely essential". For an analysis of respect for subjects and field experiments, see Teele (2014).

Scholars' beliefs about what is ethical and what is not do not carry the same moral weight as do the attitudes of uninformed, deceived, and sometimes unwilling subjects (except perhaps when we are ourselves the subjects of such studies!). But understanding scholars' beliefs is an important first step toward developing disciplinary norms and guidelines for research involving direct contact with human subjects. It also provides a valuable contrast between what our subjects think about our research, and what we think about it. In the next section, I'll discuss my survey approach to understanding attitudes about political science experimentation.

3 The survey

I conducted a survey of subjects and scholars in 2015. The survey asked respondents to read several short vignettes describing hypothetical experiments and to report their judgements on the research designs. Scholars saw four vignettes; subjects saw three. Two vignettes were identical for scholars and subjects. One was an informational field experiment, where a hypothetical researcher provided information to subjects and observed post-treatment behavior. The other vignette shared by subjects and scholars was a correspondence study field experiment, where the researcher pretends to be a client or constituent during email communication with a third party. Scholars saw two other vignettes: one described a standard laboratory experiment conducted on a college campus. The other was a foreign review scenario and described a scholar conducting an international experiment without approval from the host country. Subjects saw one additional vignette, which was a slightly different version of the foreign review case. To better identify what aspects of studies are controversial, I randomly varied features of the hypothetical studies, including consent, deception, aggregate versus disaggregate cost, the normative value of the study, debriefing, subject population type, and other features, including vignette order and the name of the researcher.

In the following paragraphs, I describe each of the hypothetical designs.

3.1 Informational Field Experiment - Both Subjects and Scholars

The informational field experiment involved a researcher sending flyers to registered voters and then observing their behavior. There were seven randomized features of the design.[15] 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, informed, and consenting.[16] 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 was also varied, capturing more or less normatively ambiguous studies. In one version, the researcher was attempting to increase flossing, presumably an unambiguously normative public health good. In a second, the researcher was trying to increase turnout and the flyer was a reminder to vote. In a third, the researcher was trying to affect vote choice and the flyer was a reminder that one candidate in an election had received a DUI conviction 5 years previously. The aggregate impact was varied, as 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 of the vignette). Finally, in some versions of the design, the scholar debriefed the subjects after the study was complete.

One version of the vignette is shown below - in this case, a study on vote choice, performed without informed consent in a close election, with debriefing, and with 1,000 subjects:

> Professor M wants to see whether a flyer has an effect on who people vote for.
>
> During an election, Professor M chooses 1,000 people from public voter registration records to be in the study. Professor M randomly divides the list of people

[15] The seven features were: informed consent, deception, size, close election, topic of study, and hypothetical researcher's name.

[16] Two features of this vignette deserve mention; both reflect my efforts to make versions of it as comparable as possible. The hypothetical dependent variable is measured through a public opinion survey, because although turnout could be observed directly in the United States, other behaviors (like flossing) cannot. In the version with consent, subjects volunteer for the study and agree to receive flyers from the researcher. The flyers include text identifying them as part of the study.

into two groups, and sends a flyer to one of the groups and nothing to the other group. The flyer reminds voters that one of the candidates was convicted of drunk driving five years ago.

The flyer is sent anonymously with no information about the study or the professor.

After the election, Professor M contacts all the people in the study and asks them who they voted for. The Professor will then see if people who received the flyer report voting differently than people who did not receive the flyer.

The election is expected to be close and the flyer might affect who wins.

The flyer has no return address and subjects are never told that they are in a research study.

The study was approved by all appropriate committees at the researcher's university, including the Institutional Review Board (IRB) or Ethics Committee.

All vignettes always contained the final sentence above - that the design had been approved by the university.

3.2 Correspondence Study Field Experiment - Both Subjects and Scholars

The second vignette described an experiment where the hypothetical researcher would interact with subjects as if s/he were a private citizen. The researcher would contact the subject, pretending to need some information, and observe whether the subject responded or not. Again, the most important treatment is consent; in some versions of the vignette, the researcher pretends to be a private citizen and subjects do not know they are in an experiment. In another version, the researcher recruits subjects, completes an informed consent process, then asks them how they would respond if someone were to ask them for information. The vignette also varies the target of the study: home sellers, businesses, or elected officials, and thus the researcher would pretend to be a potential home buyer, a potential client, or a constituent.

A number of other features of the design were randomized. The aggregate and disaggregated burden of the study were both manipulated. The number of subjects varied from from

500 to 10,000 and the burden on subjects, measured as time to provide requested information, varied from 5 to 60 minutes. I also varied the normative ambiguity of the design, which was reported as either a study on discrimination or on communication. Lastly, in some versions subjects were debriefed and could ask that their information be removed from the study; in other cases subjects were not debriefed. The text below shows the version where there is no informed consent and researcher employs deception, there are 10,000 subjects who are never debriefed, the research is on communication by elected officials, and the experiment has a high burden on subjects:

> Professor J wants to study public communication. Professor J sends an email to 10,000 elected officials. In the email, Professor J pretends to be someone who lives in the elected official's district, and asks for information about a government program.
>
> Professor J randomly varies something about the hypothetical citizen's request.
>
> Professor J will then see if elected officials are more likely to answer some types of questions than others.
>
> It takes each elected official about 60 minutes to provide the information.
>
> The elected officials are never told that they are participating in a study.
>
> The study was approved by all appropriate committees at the researcher's university, including the Institutional Review Board (IRB) or Ethics Committee.

In the version with consent, the researcher recruits volunteer subjects, shows them a hypothetical email from a constituent, customer, or home-buyer, and asks them whether or not they would answer it.

3.3 Foreign Review - Subject Version

The subject version of the foreign review focuses on the case of research in Mexico or the United States. The goal is to capture both general judgements of what subjects believe scholars should do when conducting research outside the United States, as well as perceptions of respondents as potential subjects of unapproved foreign research. In the vignette, a scholar

is conducting a public opinion survey. The scholar's home country and the country of the study are both varied between the United States and Mexico, and the time to complete a review process is varied from 90 days to two years. Thus, subjects might see a vignette asking about a Mexican researcher conducting a study in the US, a US researcher conducting a study in Mexico, or a US research conducting a study in the US, all without approval.[17] One version of the vignette follows:

> Professor K is based in the USA and will conduct a study in Mexico. The study is a public opinion survey where people are asked 10 standard questions. However, let's consider a situation where Mexican law says that all studies must be reviewed by the Mexican government for safety and for ethics, and the review takes 90 days to complete.
>
> If Professor K runs the study without completing the review process, there is almost no chance of getting caught.
>
> Professor K decides to skip the review process and just run the study.

3.4 Foreign Review - Scholar Version

The scholars' foreign review vignette examines the question of conducting research overseas without approval from the host country. In the scenario, a US-based professor is conducting research in a generic foreign country ("Country B"). There are three treatments in the vignette. The length of the review process required to obtain approval varies between 90 days, two years, and never obtaining approval. The regime is identified as a democracy or a non-democracy. The study is reported as causing risk to subjects, or no risk to subjects. One version of the vignette follows:

> Professor L is based in the US and wishes to conduct a public opinion study in a different country, which will be referred to as Country B. Country B has its own ethics review process, which is required by law in that country.

[17] I did not include the case of a Mexican researcher conducting research in Mexico, as US-based subjects' opinions on how Mexican researchers should conduct research in Mexico are not relevant to this study.

The review process in Country B is difficult to navigate and it may take up to two years to obtain permission for the study.

Although Country B is a democracy, some of the survey questions are sensitive and could cause political problems for subjects and enumerators.

Professor L decides to skip Country B's review process and proceed with the study.

The study was approved by all appropriate committees at the researcher's university, including the Institutional Review Board (IRB) or Ethics Committee.

3.5 Laboratory Experiment - Scholars Only

A fourth vignette was only shown to scholars: a laboratory experiment. These types of experiments have been used in political science for many years and are not part of the current controversy. There does exist, however, long-running disagreement between scholars on the use of deception in laboratories. Two extremes are those of economists and psychologists. Economists are generally unified in their opposition to any use of deception in laboratory experiments, because deception can pollute the population of undergraduates and make behavior in standard economic games suspect, as researchers cannot be certain of subjects' thinking.[18] Psychologists are more tolerant of deception when deemed necessary for research purposes, and when designs include debriefing after the experiment. Political science has elements of both disciplines, and this vignette was included to gauge broader opinion on such designs.

This vignette has only three variants. In one version there is no deception; in a second, there is deception after subjects have been warned; in a third, there is deception without any warning. The text of the deception with warning version is shown below:

Professor K wants to see how framing affects opinions.

[18]This opposition seems largely focused on protecting research, not with any presumed harm of deception to subjects, especially since economists are widely employing field experiments that use deception and lack informed consent.

To find out, Professor K will conduct an experiment in an empty classroom on campus. Professor K will use volunteer students as subjects to conduct the study. All subjects will sign a standard form agreeing to be in a study and will be paid for their time. The form warns volunteers that some features of the experiment may be kept hidden from them until after the study.

During the experiment, volunteers read what appear to be newspaper articles about a candidate running for city council in another state. After reading the stories, subjects are given a survey. Subjects are initially told that the story and candidate are real, although they are not. After the study, the subjects are debriefed and told that the candidates and stories were hypothetical.

The study was approved by all appropriate committees at the researchers university, including the institutional Review Board (IRB) or Ethics Committee.

3.6 Dependent Variables

For all the vignettes, subjects and scholars were asked to answer the following question:

To what extent do you agree that it is acceptable to conduct this study?

Respondents could then choose any of the following answers, which were coded from 1 (Strongly Disagree) to 7 (Strongly Agree): Strongly Disagree, Disagree, Somewhat Disagree, Neither Agree nor Disagree, Somewhat Agree, Agree, Strongly Agree.[19]

For two of the vignettes, citizen-subjects were asked an additional question. This question was designed to distinguish between subjects' general judgements about an experiment and their own feelings as potential subjects. Subjects might judge an experiment as, in principle, unacceptable, but still not care if they were included in the study. Alternatively, they might think a design acceptable but be offended if they were included. To distinguish between subjects' independent judgements of designs, and their feelings as possible subjects, I asked

[19]During trials, some respondents disliked a version of the question that specifically mentioned ethics. In the actual survey, some scholar-respondents pointed out that this question wording is potentially biased - that it would be more appropriate to ask to what extent one "agrees or disagrees", to avoid positive response bias. This question flaw should bias measured acceptability of the designs upward.

this additional question for the Informational Field Experiment and the Role Playing Field Experiment designs:

> 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?
>
> I would be glad I was in the study
>
> I would rather not have been in the study
>
> I would not care either way

The wording was slightly different for versions of the vignette where there was no deception.[20]

There were other differences between the subject and scholar surveys. The subject respondents were compensated; scholar-respondents were not. The consent form was different for the two groups. An abbreviated consent form was shown to the presumably informed scholar-respondents with a full consent available for those who were interested[21], but a full consent was shown to all citizen-subjects. I used several screening questions to remove subjects that were not paying attention.[22] As mentioned, some of the subject vignettes had an additional question about their desire to be in or be excluded from such a study. Some demographic and background questions were only relevant for one population or the other. For example, institutional affiliation type and field within political science were only asked of scholars.

3.7 Sample

The survey of scholars was conducted in two waves in the summer of 2015. The American Political Science Association generously cooperated with the study, providing a random

[20]For vignettes with consent, the question had to be modified slightly for the fully-informed version of the design. For that version, the question was, "Suppose you learned that a study like the one described above was planned for your community. Which of the following best describes how you would feel about being included in the study?", and possible answers were also slightly modified. For example, instead of "I would rather not have been in the study", respondents could choose, "I would rather not be in the study."

[21]About 5% of scholar-respondents viewed the full consent.

[22]Subjects were asked to select both "A" and "D" from a menu of five letters, A, B, C, D, and E.

sample of current and former members' email addresses in two waves.[23] The first random sample contained 5,000 email addresses which yielded 702 responses. APSA then provided an additional 9,220 addresses. In total, 1,731 of those contacted started the survey, and almost 1,600 completed the four "Agree Acceptable" questions.[24]

The survey of subjects used a panel of 3,000 respondents provided by Survey Sampling International (SSI), an online market research firm. SSI respondents are compensated for survey participation. The panel was intended to mirror the adult population of the United States. The survey was also fielded in the summer of 2015.

Table 1 compares the profile of subjects and scholars surveyed. The SSI subject sample is fairly diverse, but not perfectly representative of the United States. Homeownership among respondents is only 30%. The sample of scholars is, not surprisingly, older and whiter than the subject sample. Scholars in the study come overwhelmingly from PhD granting institutions, with only 13% from MA or BA terminal degree programs, and only 1% from community colleges. Two thirds of respondents were split almost evenly between traditional tenure-track ranks: Assistant, Associate, or Full Professors. Graduates students comprised 19%, post-docs 5%, and "Other" positions were 9%. Among scholars, all of the major fields were well-represented in the survey. Nearly half of scholar-respondents had conducted an experiment, suggesting some self-selection into the survey.

[23] In both waves, I removed email addresses of all known colleagues who helped trial the survey, discussed it with me, or had otherwise been exposed to it.

[24] In theory, this sample could yield duplicate respondents - scholars that answered the survey twice. There are two mechanisms for this. The first is that APSA's database might have multiple emails per individual, as scholars may change institutions or otherwise have different email addresses over time. The second has to do with a sampling error. APSA reported that the addresses in the two samples were unique, but when the survey was administered, some potential respondents informed me that they had already taken the survey. After some investigation, I discovered that there were 2,862 duplicate email addresses in the two samples. To avoid any subjects' taking the survey twice, I immediately contacted duplicate respondents and asked them not to take the survey. However, because the survey was fully anonymized, I cannot link any email addresses to responses, and thus cannot identify which subjects have completed the survey and which have not. Thus, in theory, some respondents could have completed the survey twice, although given that subjects were uncompensated, scholars would have to feel very strongly to bother doing so.

4 Results

4.1 Informational Field Experiment

Figure 1 shows the impact of informed consent and research topic on attitudes about informational field experiments. The left panels show results for subjects; the right panels show results for scholars. In each of the graphs in the first row, 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 (1-7 scale). The points show the mean acceptability, with 95% confidence intervals. 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 in the graphs. First, both subjects and scholars are sensitive to the presence or absence of consent; surreptitious experiments lower acceptability. 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 content of the experiment, in particular the normative ambiguity of the design. I expected highest acceptability for the flossing treatment, followed by the GOTV and the DUI treatments. For subjects, the GOTV treatment is the most acceptable, followed by the flossing treatment, and then the DUI treatment.[25] For scholars, the expected trend is observed: acceptability is highest for the flossing treatment, slightly lower for the GOTV treatment, and lowest for the DUI treatment.

Third, although most of the trends are the same for scholars and subjects, scholars appear to be more sensitive than subjects to both the type of study and the presence of informed consent. On average, scholars are almost twice as responsive to treatments than are subjects. For scholars, the mean difference in acceptability between designs with and without informed consent is 2.10; for subjects, the difference is 1.22. Looking just at studies without informed

[25] Perhaps this reflects a general distaste for dentists or dental care.

consent, for scholars, agreement is 1.53 higher for Flossing reminders than for DUI reminders; for subjects, the difference is .67.

A look at the underlying distribution of responses is helpful here, shown in Figure 2. The barplots show the distribution of responses for the informational field experiment, by study population (scholars or subjects) and by informed consent. For scholars, the contrast between designs with and without consent is stark. The modal response when using consent "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 distribution inverts, and the most common responses are "2" and "1" - disagreement that the design is acceptable. For all cases, scholars have opinions on these issues: only about 5% of respondents choose the "neither agree nor disagree" response.

The distribution of subject responses shows a similar shift, but is 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%. The modal response for both subjects, with and without consent, is "2" - "Agree Acceptable". 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.

Table 2 shows results from regressions of treatments on reported acceptability.[26] The model includes indicator variables for the treatments (DUI and GOTV), with Flossing as the excluded category. The results mirror findings from the graph, above: both types of respondents respond negatively to designs without consent. For scholars, the Flossing design (excluded) is the most acceptable, followed by the GOTV design and then the DUI design. Subjects are less favorable to flossing than the GOTV treatment, but like scholars have weakest support for the DUI design. Again, subjects are generally less responsive to the designs - estimated coefficients are smaller and the model fit is weaker.

The models also show the impact of the other treatments. Both subjects and scholars reactive negatively to explicit deception - sending a flyer that is attributed to a fake organization significantly lowers mean acceptability ($-.382$ for subjects, $-.299$ for scholars). Other variables yield mixed results. Scholars are concerned about affecting elections - running an

[26]Full models with controls available in online appendix.

experiment that could affect an electoral outcome reduces acceptability ($-.577$) - but subjects had no significant response to the same treatment. Regarding the aggregate impact of the study, there is not strong evidence that subjects or scholars are concerned about the size of the experiment. Although the estimated coefficient on size is negative for both groups, the estimates are not significantly different from zero. 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.

Additional models with more controls (not shown) do not change any of the core results. For subjects, more educated respondents are significantly more likely to find designs acceptable. Older respondents and female respondents were less likely to find designs acceptable. Other controls did not yield consistently significant coefficients. For scholars, field, race, gender, institution type, and position had no impact on acceptability. Only one control variable was significant: "Ever Experiment" - an indicator variable for whether the respondent had ever conducted an experiment - had a positive and significant coefficient, indicating that experimentalists are generally more accepting of these designs than non-experimentalists.

The lower-left graph in Figure 1 shows subjects' responses to the second question - how they would feel about participating in such a study. This question was designed to encourage respondents to think as subjects, rather than abstractly consider the ethics of any particular vignette from a third-party perspective. Again, the x-axis is the type of treatment (Floss, GOTV, or DUI/negative campaign message). The y-axis measures the proportion of respondents who answered,"I would rather not have been in the study." As previously, the dashed line shows mean responses for designs without consent, and the solid line shows responses for designs with consent.

For the cases with informed consent, few respondents wish to avoid the GOTV or Flossing treatments - just 14% and 16%, respectively, reported that they would rather not participate. For the DUI case, rejection rose considerably, with 30% reporting wanting to avoid the study. Logistic regression on an indicator variable for a preference not to participate is shown in

Table 3, with similar results.

Designs without informed consent had a much higher rejection rate. For the Flossing study, the rejection rate doubled to 29%. The GOTV study saw rejection increase slightly, to 20%. And even in a minimally intrusive design that many subjects judged acceptable, almost half (46%) would rather not have been in the DUI experiment.[27]

4.2 Correspondence Study Field Experiment

Figure 3 shows results for the Correspondence Study Field Experiments, using the same graph format as in the previous example. Here, however, the x-axis is categorical and reports the target of the study: Elected Official, Business owner, or Home Seller.

The primary result here is again that there is a strong reaction against designs without consent among scholars, and a smaller one - though still significant - among 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, mean agreement falls, by 1.82 to 4.66. Subject responses echo those of scholars, but with smaller differences between designs with and without consent (an mean difference of .72).

The second finding here is that there is only a modest impact of the target subject

[27]Subjects' optional open-ended comments reflect all the core debates. Consider some quotes from the Flossing vignette: R#130 writes,"This flyer is not only for research, but could be considered a public service. I see nothing harmful or invasive in sending this out.". "The lie about the health organization does bother me a little bit. But ... flossing is good." R#206. Other respondents reacted more strongly to the deception of a fake group sending the flyer: "dont lie!!!! ... Legally he may be allowed to do it but morally and ethically it is wrong and I am mindful that the university knowing this allowed this survey to be send using university funds – wrong wrong". A number of respondents mentioned some form of consent - "People should be informed that they are in a study and as to how their results will be used.R#159".

There was more suspicion regarding the studies involving elections: "it's unethical to send negative information about a candidate" R#795" and "It appears to be an effort to possibly influence the election under the guise of a research study", R#840. Some respondents expressed concern about broader implications, for example: "I believe if people should know that they will be part of a research study. It is in my opinion unethical to lie to them. If the professor lied about this then the next time he might want to push the boundary of his research study that might go to far that could affect the subjects lives." R#370

population on acceptability. Although elected (public) officials are exempt under the current version of the Common Rule, treating them is less popular than treating business owners. As expected, home sellers are the least acceptable target for such studies, though the difference between populations is modest for both subjects and scholars.

When adding controls and running regressions (Table 4), these results persist. An indicator variable for "No Consent" has a strong negative impact on responses. Comparing targets, business owners are the most acceptable, followed by elected officials and then by home sellers, although government officials are only significantly more acceptable for the subject sample. As with informational field experiments, the apparent normative value of the topic is relevant - designs that study discrimination are significantly more acceptable than those that study communication, customer service, or constituency service. The estimated coefficient on an indicator variable for a study topic of discrimination was approximately .30 for both populations. However, this higher tolerance for normatively valuable studies does not offset the larger cost of skipping consent - the estimated increase in acceptability associated with studying discrimination was .30, but the coefficient on "No Consent" was -.70 for subjects and -1.8 for scholars. Lastly, a higher burden on subjects reduces acceptability for both groups. The size of the study and debriefing are irrelevant for both samples.

The last two columns report models where design features are interacted with consent. For subjects, *No Consent* interacts with the target of the study, echoing Figure 3. Subjects find role-playing designs without informed consent least acceptable when deployed on home sellers, slightly more acceptable when used on government officials, and most acceptable when used on businesses.[28] For scholars, there are larger negative reactions to deception, and smaller, though still significant, interactions between consent and subject type. A lack of consent lowers mean acceptability by 2.11 when targeting Homeowners, but only by 1.69 when targeting government officials (not significant) and by 1.57 when targeting Business Owners.

Finally, the lower-left panel in Figure 3, and the logistic regressions in Table 5 investigate the determinants of subjects' feelings about being in such a study. In the graph, the

[28]The interaction of *Deception* and *Govt Offcl* is +.40 and of *Deception* and *Business Owner* is +.74, showing more acceptability than the excluded case of a study using deception and targeting homeowners.

proportion of subjects preferring not to participate in such a design, by target and deception. In this case, most respondents could only reasonably imagine being a homeowner or business owner - not an elected official. Consequently, the follow-up question was only asked for the homeowner and business versions of the vignette. As with the informational field experiment, rejection is very low in the case of informed consent and in this case, does not vary with target. For the Business version of the design, only 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 potential customer or potential home-buyer, rejection is much higher, at 28% and 41%, respectively. The logistic regressions, again with controls not shown, largely reiterate these points.

Results here echo findings from the previous 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, significant numbers of subjects would rather not be included in studies without consent.

4.3 Subjects Only: Foreign Review

I now examine the question of foreign review of international experiments. In all the vignettes, respondents judge a study where a scholar chose not to comply with review procedures. Figure 4 shows mean acceptability for subjects for their version of the foreign review vignette. In this design, acceptability of *any* experiment without approval was low, but varied across scenarios. Acceptability was highest for US-based scholars conducting research within the US. It was lower for US conducting research without approval in Mexico, and lowest for Mexicans conducting research in US. Mean acceptability , and variance, were low, and all scenarios had mean acceptability below 4.0. Approval was slightly higher in cases where approval of the experiment would take two years, instead of ninety days, but this difference was small and insignificant. Notably, subjects were slightly more tolerant of skipping review here in the United States than doing so in other countries. In practice, scholars appear more likely to comply with US research rules than with those in other countries.

4.4 Scholars Only: Foreign Review and Laboratory Experiments

Figure 4 also shows scholars' attitudes toward skipping foreign review, and toward laboratory experiments. Although many US-based scholars conducting experiments overseas are entirely skipping foreign review, scholars generally are opposed to such designs. The average acceptability score for treatments reported in Figure 4 are all well below 4 on the 1-7 scale, indicating that the average respondent is opposed to skipping foreign review. In regressions, these results persist with controls: time to review increases acceptability, and when Country B is a democracy, acceptability is reduced (See Table 6). Surprisingly, whether or not the survey posed risk to subjects was not significant.[29] In all regression models, the intercept is low and all estimated coefficients are small - so all scenarios have low acceptability.

In my own experience, among scholars of comparative politics, there is a general sense that asking for foreign review is a waste of time - I've been called "stupid" by fellow comparitivists for even trying to navigate Brazil's difficult ethical review procedures. For those conducting research in authoritarian countries, ever obtaining official approval from the host government is extremely unlikely. Consequently, many scholars conducting research in other countries skip local review. But clearly the rest of the field does not find this strategy acceptable.

A very different pattern emerges in the Laboratory Experiment Design. Mean Acceptability is very high for all scenarios - ranging from 6.48 to 5.92 on the 1-7 Acceptability scale. Acceptability fell as deception increased, but only modestly. Criticism of this design was sometimes based on deception, but sometimes based on the triviality of the research design.[30]

5 Discussion

I conclude by briefly summarizing the findings, discussing some of the weaknesses of this study, and examining the implications of results for the discipline.

[29] This may be a floor effect - respondents already strongly oppose such designs, so risk could have little impact on their already largely negative judgements.

[30] Consider the following two comments on the laboratory experiment: R# 162 "The design contaminates the subject pool. If people can't believe the instructions they receive how can we believe the information they provide...."., versus R# 244"...My primary objection, however, is that the research is trivial."

For field experiments, there are three general 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 reduced mean acceptability scores in every analysis, and differences were always statistically significant. Scholars were very sensitive to informed consent. For designs with consenting subjects, nearly all academics agreed, or strongly agreed, that these are acceptable. Designs without informed consent were polarizing, with 36% agreeing to some extent that such designs are acceptable, in contrast with 58% expressing some degree of disagreement. In both cases, very few scholars were ambivalent; "Neither Agree nor Disagree" responses were below 6%.

Subjects' reactions to features of the vignette were less dramatic, but had the same general trend: negative response to deception and to studies without some form of consent. For both subjects and scholars, debriefing did not assuage these reactions; the impact of debriefing on acceptability was small and insignificant.

Second, other features of the design matter, especially the topic and treatment. Studies with clear normative value are more acceptable that those with normatively ambiguous topics. Informational Field Experiments promoting flossing or turnout are judged more acceptable than negative campaign messages. Correspondence studies on discrimination had higher acceptability than those on customer service, constituency service or communication. These results held for both subjects and scholars. Thus, while respondents rejected deception, they were also slightly more tolerant when the research could clearly contribute to the social good. The potential aggregate impact of designs yielded mixed findings. Neither subjects nor scholars distinguished between small and large experiments. Scholars were very responsive to the possibility of affecting an election - this significantly reduced acceptability; subjects had a similar, but not significant, response.

Third, subjects were generally less responsive to all treatments, and on average lukewarm toward many of them, while scholars reacted strongly to small design changes. As a result, subject opinion moved in a narrow band, while scholars' opinions often jumped sharply above and below subjects in response to design changes. I had hypothesized that subjects would generally be more critical of our designs than we are. Instead, scholars were more critical of designs without consent, and more accepting of designs with consent.

There are two possible interpretations of this result. One possibility is that subjects have generally lukewarm and positive views of our research, and simply don't care as much as we do about issues of consent, deception, aggregate impact, and type of study. Perhaps we should not worry so much about these issues. Alternatively, subject-respondents might be less responsive for other reasons. As compensated respondents, they may have simply paid less attention, and as non-academics with presumably little training in research design or ethics, they may not have understood the vignettes or the issues involved.

Lastly, perhaps the most important takeaway is that large percentages of subjects in field experiments conducted without informed consent would rather not be in such studies. Opposition was lowest in cases where deception was minimized, where burden on subjects was minimized, and where the topic had some clear normative value - investigating discrimination, or promoting turnout or flossing. Even in these cases, over 20% consistently reported preferring not to be included. On the other hand, designs with ambiguous value (negative advertising or communication), with additional deception (flyers attributed to non-existent groups), were widely opposed. In some cases, nearly than 50% of subjects reported that they would rather not be in such studies. It is difficult to imagine a defense of research where nearly half of subjects are effectively forced into participating against their will.

I offer some general suggestions for field experiments. First, use some form of consent and avoid deception whenever possible. If traditional informed consent is not an option, consider some of the alternatives offered in Humphreys (2014). If you do run a field experiment without any consent, consider four strategies to assuage criticism, proposed in Desposato (2016*a*): 1) Do good - make sure that treatments and dependent variables and other side effects have clear normative value, as much as possible; 2) Tread Lightly - minimize burden, spillovers and sample size. If studying an election, pick one where you cannot affect the result, use balanced sample, and minimize sample size; 3) Confess - debrief all, or a sample of your subjects and affected bystanders[31]; 4) Compensate - pay your subjects a standard and non-trivial amount for their time. The first two suggestions were directly supported by findings in this paper - subjects responded positively to normatively valuable and less burdensome designs. Regarding debriefing and compensation, debriefing did not significantly increase

[31]Technically, debriefing this is required when praticaly possible by the Common Rule.

acceptability in this study, and compensation was not tested. However, as I have argued elsewhere, both show respect for subjects, encourage learning about subjects' opinions of our designs, and impose natural constraints on our ambition(Desposato, 2016*a*).[32]

Regarding foreign review of international experiments, there is widespread disapproval of experiments conducted overseas without any local review. Most countries have regulations governing research and most scholars conducting experiments overseas ignore those regulations. Among scholars, there were modest concessions to lengthy review processes or and regime type, but disapproval remained high in all cases. For subjects, who saw a slightly different scenario, approval was low whether the research was conducted by a US researcher in Mexico, a Mexican researcher in the US, or a US researcher in the US. Indeed, for subjects, the most acceptable case where a scholar might skip review was that of a US research conducting an experiment in the US - which is exactly the case where full compliance with rules is most likely to be the norm.

This result suggests a great need for discussion in the field, given the wide gap between attitudes of subjects and scholars on the one hand, and actual practice by those conducting experiments overseas on the other. Full compliance with foreign research regulations would severely constrain research agendas in some countries, end research in many others, and greatly increase the length of time needed to complete a dissertation or other project in comparative politics. It is also important to note that many of these research rules apply to all studies, observational or experimental, qualitative or quantitative. For some scholars, this would imply a multi-year approval process just to observe an election and interview some politicians. This standard seems excessive, and may reflect the limitations of the vignettes I presented. Responses might have shown more flexibility if the full impact of compliance with

[32]An interesting approach to informed consent is discussed by Koenig (2014). Ethics researchers developing consent procedures for genetic testing confronted the possibility that full counseling for an individual could require days of consent procedures. As an alternative, they convened a representative group of citizens for a four-day discussion of appropriate procedures, including recruitment, unexpected findings, and best practices. In theory, if the panel approves of the research, then actual subjects might see a much abbreviated informed consent. The paradigm is one of submission not to a procedures, but to a "governance scheme"(Koenig, 2014, p. 34) (Humphreys (2014) calls this "proxy consent".). For political science field experiments, such an approach could yield much deeper understanding of subjects' views than this survey experiment.

sometimes corrupt rules had been explained, or if alternative forms of local review had been offered, including evaluation of the contextual appropriateness by trusted local scholars. I also did not ask about other strategies for dealing with impossible or corrupt foreign review processes, for example, collaborations with local scholars or internet based studies using offshore servers (not uncommon for opinion research in authoritarian regimes that would never approve some basic surveys). Either way, these issues also clearly need discussion.

There are many limitations to the findings and conclusions herein. Neither of the two samples is a perfect random draw from the populations of interest, and observed differences between samples could reflect sampling frames. Subject responses came from a paid panel, and scholars represent the subset of APSA members who chose to respond to my survey invitation. Results might reflect specific features of the United States, and findings might be reversed in other contexts. For example, considering the informational field experiment, It is easy to imagine that subjects in a different country might be grateful to receive negative advertisements, if their normal campaign environment never provides information about candidates' flaws. In other settings, subjects might have generally higher tolerance for research and greater trust in scholars, and find these minimal risk designs completely acceptable. There are also many additional variables that deserve investigation. Field experiments without consent might be welcomed if, post-study, subjects all received an unexpected cash payment. Opposition to any specific designs might easily be assuaged if scholars could explain the aims and importance of the research, and the reasons for the chosen approaches. At the same time, rejection of these designs could be much higher for more aggressive studies.

For all these reasons, this study should not be seen as the last word, but merely as some introductory remarks in an overdue conversation that needs broad participation. Looking forward, some easy first steps could involve debriefing subjects (especially in studies that lack any consent) and reporting their opinions, testing whether respondents are more tolerant of our research when given more information about our aims and methods, and extending this study to other populations. More generally, discussions of research ethics should become part of the research enterprise - they should be discussed in publications, reviews and presentations, not just in post-panel lobby conversations.

References

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.".
URL: *http://oxfordmedicine.com/view/10.1093/med/9780199297368.001.0001/med-9780199297368-chapter-004*

Bositis, David A. and Douglas Steinel. 1987. "A Synoptic History and Typology of Experimental Research in Political Science." *Political Behavior* 9(3):263–284.

Broockman, David E. 2013. "Black Politicians Are More Intrinsically Motivated to Advance Blacks' Interests: A Field Experiment Manipulating Political Incentives." *American Journal of Political Science* 57(3):521–536.
URL: *http://www.jstor.org/stable/23496636*

Cohen, Rick. 2014. "Hewlett-Funded Political Science Experiment Creates Political Fuss in Montana." *Nonprofit Quarterly* .
URL: *https://nonprofitquarterly.org/2014/10/28/hewlett-funded-political-science-experiment-creates-political-fuss-in-montana/*

Declaration of Helsinki. 2008 [1964].
URL: *http://www.wma.net/en/30publications/10policies/b3/17c.pdf*

Desposato, Scott. 2016*a*. Conclusion and Recommendations. In *Ethics and Experiments: Problems and Solutions for Social Scientists and Policy Professionals*, ed. Scott Desposato. Routledge Series in Experimental Political Science Routledge pp. 267–289.

Desposato, Scott. 2016*b*. Introduction. In *Ethics and Experiments: Problems and Solutions for Social Scientists and Policy Professionals*, ed. Scott Desposato. Routledge Series in Experimental Political Science Routledge pp. 1–22.

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):627–635.

Findley, Michael G., Daniel L. Nielson and J. C. Sharman. 2014. *Global Shell Games: Experiments in Transnational Relations, Crime, and Terrorism*. Cambridge University Press.

Gelman, Andrew. 2010. "$63,000 worth of abusive research . . . or just a really stupid waste of time?".
URL: *http://andrewgelman.com/2010/05/06/63000$_w$orth$_o$f/*

Green, Donald P. 2004. "Mobilizing African-American Voters Using Direct Mail and Commercial Phone Banks: A Field Experiment." *Political Research Quarterly* 57(2):245–255.
URL: *http://www.jstor.org/stable/3219868*

Grose, Christian R. 2016. Field Experiments on Elected and Public Officials: Ethical Obligations and Requirements. In *Ethics and Experiments: Problems and Solutions for Social Scientists and Policy Professionals*, ed. Scott Desposato. Routledge Series in Experimental Political Science Routledge pp. 227–238.

Humphreys, Macartan. 2014. "How to make field experiments more ethical.".

Koenig, Barbara. 2014. "Have we asked too much of Consent?" *The Hastings Center Report* 44(4):33–34.

Lieberman, Evan, Daniel N. Posner and Lily Tsai. 2014. "Does Information Lead to More Active Citizenship? Evidence form an Education Intervention in Rural Kenya." *World Development* 60:69–83.

Loewen, Peter John and Daniel Rubenson. 2011. "For want of a nail: Negative persuasion in a party leadership race." *Party Politics* 17(1):45–65.
URL: *http://ppq.sagepub.com/content/17/1/45.abstract*

Malesky, Edmund J. 2016. Manipulating Elites. In *Ethics and Experiments: Problems and Solutions for Social Scientists and Policy Professionals*, ed. Scott Desposato. Routledge Series in Experimental Political Science Routledge pp. 217–226.

McDermott, Rose. 2002. "Experimental Methods in Political Science." *Annual Review of Political Science* 5:31–61.

Mendez, Matthew S. and Christian R. Grose. 2014. "Revealing Discriminatory Intent: Legislator Preferences, Voter Identification, and Responsiveness Bias." *USC CLASS Research Paper No. 14-17* .

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.

National Institute on Money in State Politics. 2014. "Show me contributions to candidates (Standard elections) in elections in Montana 2014.".

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 Routledge pp. 198–216.

O, Ana L. De La. 2013. "Do Conditional Cash Transfers Affect Electoral Behavior? Evidence from a Randomized Experiment in Mexico." *American Journal of Political Science* 57(1):1–14.
URL: *http://www.jstor.org/stable/23496539*

Olken, Benjamin A. 2010. "Direct Democracy and Local Public Goods: Evidence from a Field Experiment in Indonesia." *The American Political Science Review* 104(2):243–267.
URL: *http://www.jstor.org/stable/40863719*

Roberts, Laura Weiss, Teddy D. Warner, Janet L. Brody, Brian Roberts, John Lauriello and Constantine Lyketsos. 2002. "Patient and psychiatrist ratings of hypothetical schizophrenia research protocols." *American Journal of Psychiatry* 159:573–584.

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* p. 115.

The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. 1979. The Belmont Report: Ethical Principles and Guideline for the Protection of Human Subjects of Research Office. Technical report U.S. Government Printing Office.

The Nuremberg Code. 1947.

Tran, Anh. 2015. "Ethical issues in conducting political science research in authoritarian regimes." *Working Paper* .

U.S. Department of Health and Human Services. 2015. "Federal Policy for the Protection of Human Subjects, http://www.hhs.gov/ohrp/humansubjects/commonrule/.". Accessed on April 1.

Table 1: Descriptive Statistics for All Respondents

	Subjects		Scholars	
Gender				
Male	48 %	1443	65 %	1443
Female	52 %	1558	35 %	1558
Age				
18-25	12 %	399	2 %	43
26-35	17 %	575	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
Home				
Not a Homeowner	62 %	2110	56 %	966
Homeowner	26 %	891	33 %	564
No Response	11 %	384	12 %	201
Race				
Native American	1.7 %	58	0.3 %	5
White	72.9 %	2467	76.7 %	1328
Black	7.4 %	252	1.3 %	23
Latino	6.3 %	213	5.6 %	97
Asian	3.1 %	105	2.8 %	49
Other / No Response	0.9 %	32	10.1 %	174

Table 1: Descriptive Statistics for All Respondents (continued)

	Subjects		Scholars	
Education				
Less Than High School	2 %	57		
High School	19 %	570		
Some College	29 %	885		
College	35 %	1036		
Grad Degree	15 %	453		
Position				
Grad Student			19 %	291
Post-Doc			5 %	77
Assist Prof			24 %	376
Assoc Prof			21 %	320
Full Prof			22 %	348
Other			9 %	137
Institution				
N/A	100	3385	11 %	186
Phd			72 %	1249
MA			7 %	128
BA			6 %	96
Comm. College			1 %	14
Other			3 %	58

Table 1: Descriptive Statistics for All Respondents (continued)

	Subjects	Scholars
Location		
Based in US	81 %	762
Based outside of USA	19 %	175
Field(s)		
American	35 %	614
Comparative	38 %	652
IR	20 %	353
Theory	9 %	157
Methods	14 %	245
Other	10 %	174
Ever Experiment?		
No	48 %	745
Yes	52 %	808

Table 2: Acceptability of Informational Field Experiments

	Subjects	Scholars	Subjects	Scholars
No Consent	-0.614*	-1.558*	-1.019*	-2.075*
	(0.075)	(0.131)	(0.140)	(0.215)
GOTV	0.087	-0.314*	-0.334*	-0.549*
	(0.075)	(0.121)	(0.106)	(0.153)
DUI	-0.878*	-1.871*	-1.100*	-2.234*
	(0.075)	(0.112)	(0.105)	(0.160)
Fake Group	-0.382*	-0.299*	-0.388*	-0.340*
	(0.087)	(0.145)	(0.086)	(0.151)
100k Subjects	0.041	-0.031	0.038	-0.129
	(0.061)	(0.094)	(0.086)	(0.127)
Close Election	-0.072	-0.557*	-0.049	-0.594*
	(0.061)	(0.094)	(0.087)	(0.127)
No Consent*GOTV			0.836*	0.524*
			(0.150)	(0.253)
No Consent*DUI			0.448*	0.706*
			(0.149)	(0.223)
No Consent*100K			0.008	0.214
			(0.122)	(0.189)
No Consent*Close			-0.038	0.086
			(0.122)	(0.188)
Constant	5.551*	6.313*	5.751*	6.578*
	(0.074)	(0.114)	(0.095)	(0.143)
R-squared	0.114	0.333	0.122	0.337
N	3023	1464	3023	1464

Table 3: Would Rather Not Participate in an Informational Field Experiment (Subjects Only)

	Model 1	Model 2
No Consent	0.480*	0.919*
	(0.113)	(0.221)
GOTV	-0.226	0.166
	(0.120)	(0.190)
DUI	0.823*	1.024*
	(0.108)	(0.170)
Fake Group	0.345*	0.356*
	(0.121)	(0.122)
100k Subjects	-0.066	-0.015
	(0.091)	(0.139)
Close Election	0.015	0.104
	(0.091)	(0.139)
No Consent*GOTV		-0.657*
		(0.246)
No Consent*DUI		-0.332
		(0.221)
No Consent*100K		-0.092
		(0.185)
No Consent*Close		-0.163
		(0.185)
Constant	-1.631*	-1.896*
	(0.116)	(0.167)
N	2678.000	2678.000

Table 4: Acceptability of Correspondence Study Field Experiments

	Subjects	Scholars	Subjects	Scholars
Target: Businessowner	0.403*	0.312*	-0.078	-0.029
	(0.073)	(0.108)	(0.123)	(0.185)
Target: Govt Official	0.199*	-0.007	-0.068	-0.288
	(0.072)	(0.108)	(0.126)	(0.188)
No Consent	-0.691*	-1.838*	-1.144*	-2.107*
	(0.072)	(0.108)	(0.154)	(0.230)
Studying Discrimination	0.297*	0.300*	0.165	0.018
	(0.059)	(0.088)	(0.102)	(0.153)
Number of Subjects	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Burden on Subjects	-0.004*	-0.011*	-0.004	-0.002
	(0.001)	(0.002)	(0.003)	(0.004)
Debrief (if deception)	-0.062	0.055	-0.060	0.064
	(0.072)	(0.108)	(0.072)	(0.107)
No Consent * Biz Owners			0.735*	0.534*
			(0.152)	(0.227)
No Consent * Govt Offcl			0.403*	0.420
			(0.154)	(0.229)
No Consent * Discrimination			0.197	0.438*
			(0.125)	(0.187)
No Consent * Size			-0.000	0.000
			(0.000)	(0.000)
No Consent * Length			-0.000	-0.014*
			(0.003)	(0.005)
Constant	5.160*	6.279*	5.453*	6.449*
	(0.083)	(0.122)	(0.121)	(0.184)
R-squared	0.059	0.207	0.066	0.215
N	3026	1597	3026	1597

Table 5: Would Rather Not Participate in a Correspondence Study Field Experiment (Subjects Only)

	Model 1	Model 2
Would Rather Not Participate		
No Consent	0.735*	1.059*
	(0.136)	(0.269)
Target: Businessowner	-0.469*	-0.161
	(0.106)	(0.206)
Studying Discrimination	-0.310*	-0.209
	(0.106)	(0.207)
Number of Subjects	0.000	-0.000
	(0.000)	(0.000)
Burden on Subjects	0.007*	0.009
	(0.003)	(0.005)
Debrief (if deception)	0.182	0.177
	(0.123)	(0.123)
No Consent * Biz Owners		-0.417
		(0.240)
No Consent * Discrimination		-0.173
		(0.241)
No Consent * Size		0.000
		(0.000)
No Consent * Length		-0.004
		(0.006)
Constant	-1.247*	-1.500*
	(0.172)	(0.223)
N	1819.000	1832.000

Table 6: Acceptability of Experiments - Other Scenarios

	Subjects Local Review	Scholars Local Review	Scholars Lab
American Prof in Mexico	-0.178*		
	(0.084)		
Mexican Prof in USA	-0.306***		
	(0.083)		
Time for Review	0.136*	0.297***	
	(0.068)	(0.054)	
Country B Democracy		-0.436***	
		(0.089)	
Risk to Subjects		-0.157	
		(0.089)	
Warn Subjects			0.153
			(0.082)
Deception			-0.561***
			(0.082)
Constant	3.094***	2.763***	6.483***
	(0.068)	(0.092)	(0.058)
N	3032	1597	1594
r2	0.006	0.033	0.031

Figure 1: Atittudes Toward Informational Field Experiments

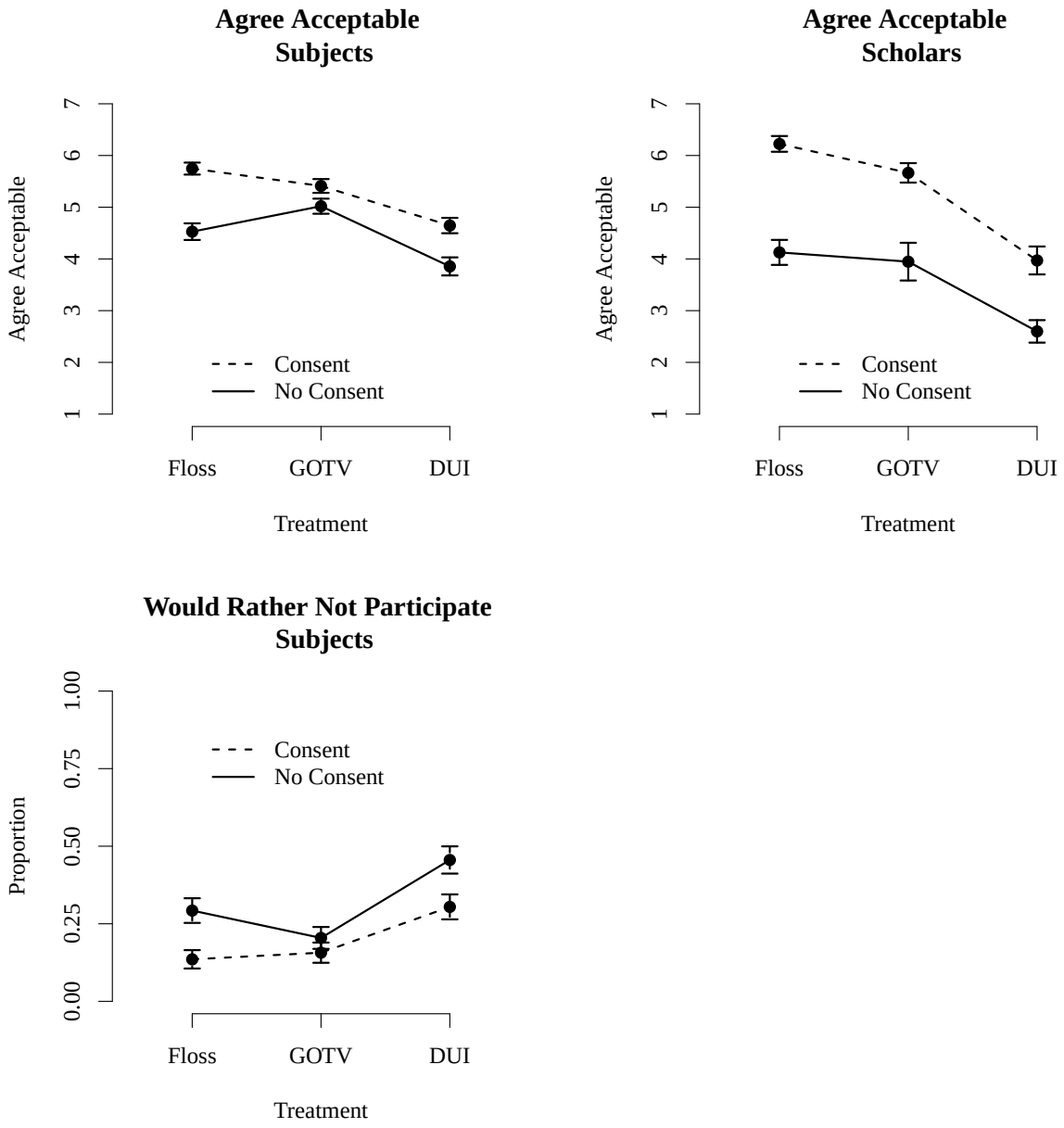

Figure 2: Acceptability Informational Field Experiments

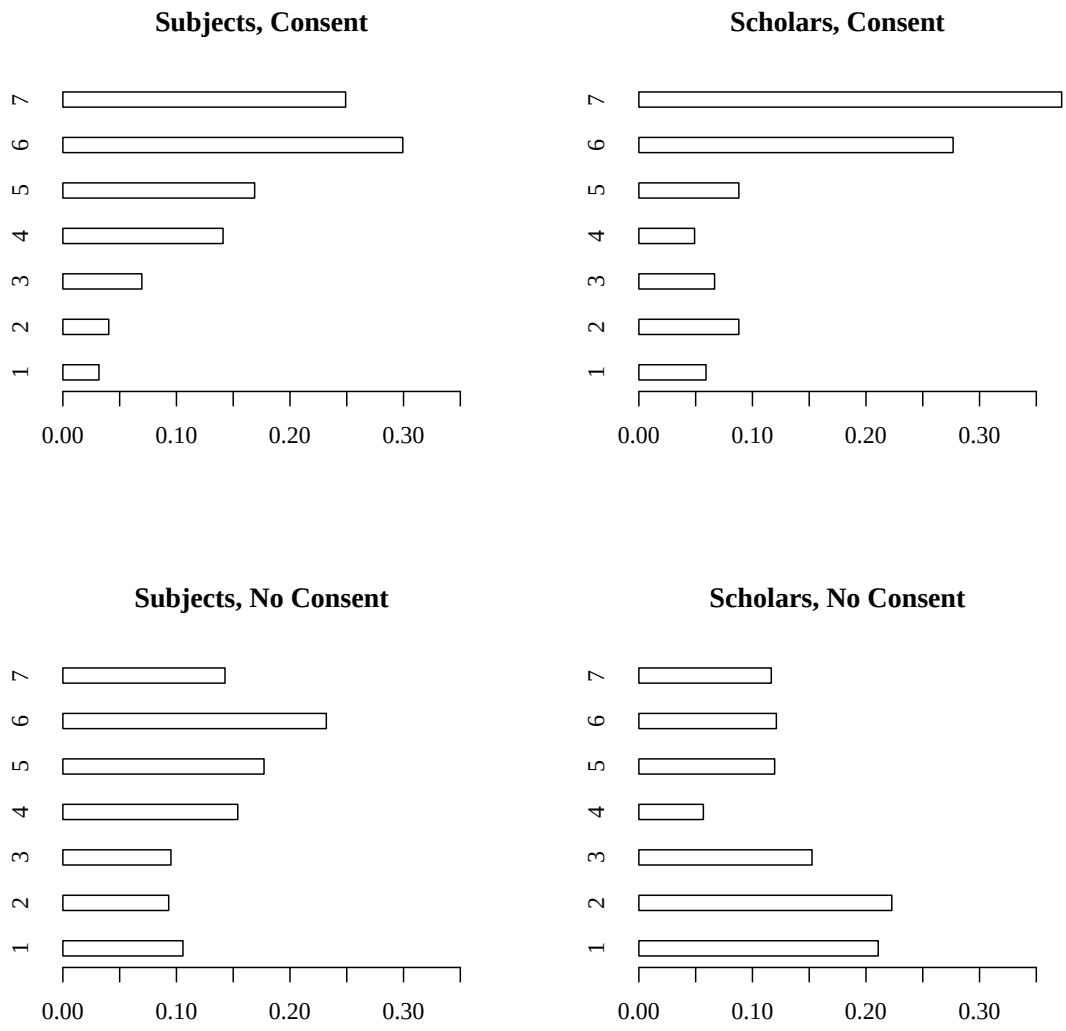

Figure 3: Attitudes Toward Correspondence Study Field Experiments

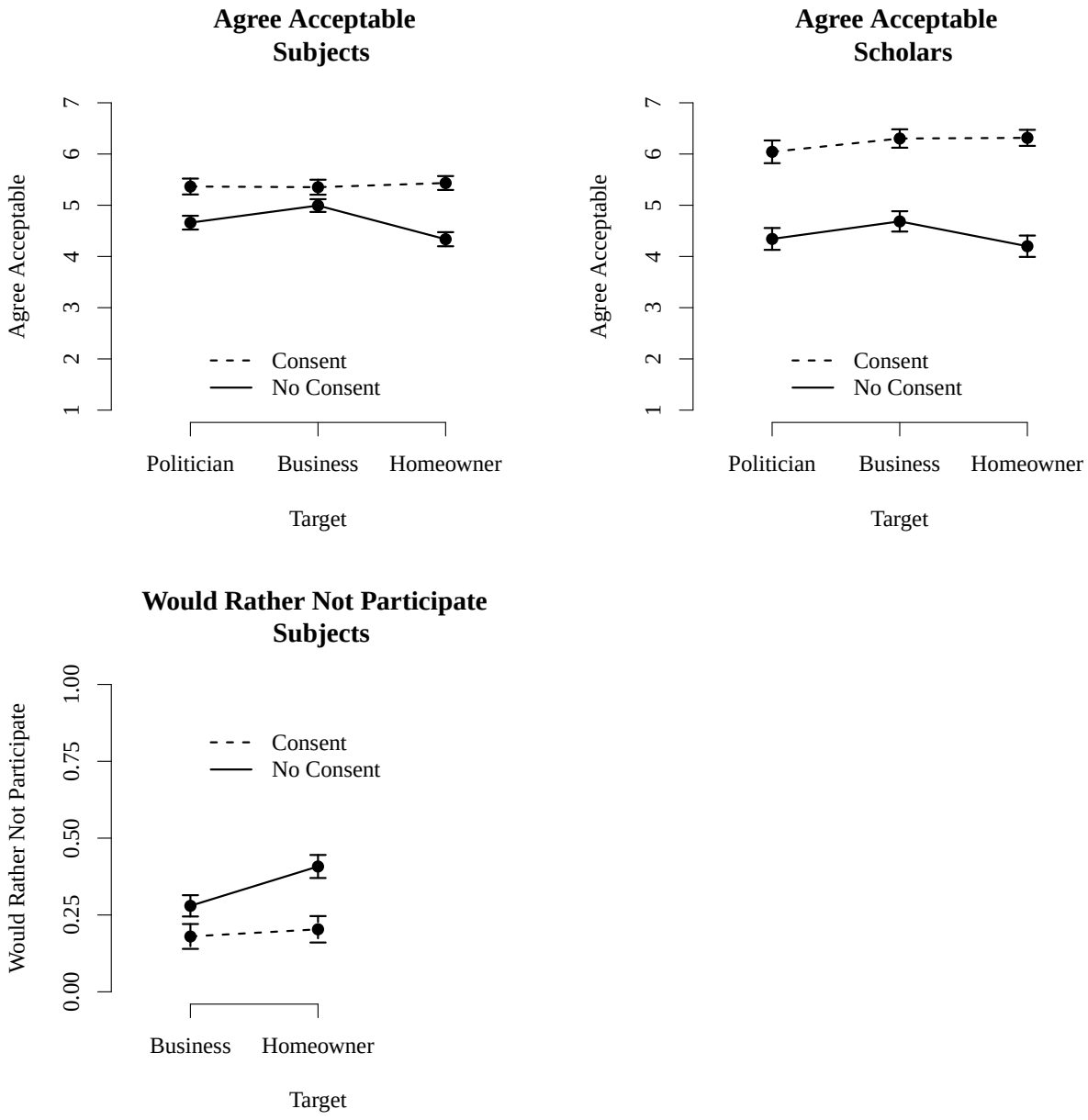

Figure 4: Attitudes Toward Foreign Review and Lab Experiments