

# Ethical Challenges and Some Solutions for Field Experiments\*

Scott Desposato  
UZH and UCSD  
swd@ipz.uzh.ch

November 3, 2014

## **Abstract**

The recent controversy over a field experiment conducted in Montana during an election has many political scientists debating the ethics of interventions in “the real world”. Much of this discussion focuses on the fact that the experiment may have violated electoral law and may not have had all required IRB reviews. However, absent some technical shortcomings, the study is identical to dozens of others that have been run all over the world. The bigger questions here are whether we can ethically run experiments that could affect millions of subjects and bystanders without their consent. I discuss some of the features that distinguish political science from other fields and offer some suggestions for best practices in field experiments.

---

\*Several paragraphs in this paper were used in a very similar form as a short post on TMC .

“What would happen if someone did this study in the United States” This was a discussion question at the end of a UCSD conference on ethics in comparative politics experiments last year (NSF#1251510, proceedings to be published in 2015 by the Routledge Studies in Experimental Political Science). In many developing countries, experimental research seems like a Wild West where almost anything goes. One way we might gauge the appropriateness of an experiment conducted overseas would be to ask how it might be received were it run in the United States. Well, now we know.

Since the controversy erupted over a field experiment in Montana (Henceforth, “The Montana Study”), the Secretary of State of Montana has filed a complaint, the presidents of Stanford and Dartmouth have apologized, and the study has been widely critiqued by both academics and journalists.

The PIs ran a standard electoral field experiment in Montana, California, and New Hampshire. They provided ideological information about nonpartisan judicial candidates by placing them on a left-right scale anchored by President Obama and Mitt Romney. The mailer included the state seals in Montana and in California, which is prohibited without approval. Including the seal and the title Official Voter Guide gave it the appearance of a state document. The study was approved by the Dartmouth IRB, though it may not have been reviewed at Stanford.

Ironically, although this experiment has proved controversial, it is typical of dozens of experiments that have been run all over the world during elections. Some of these have also been illegal and violated election laws, and some have not had required human subjects approval. To their credit, the PI’s of the Montana Study identified their institutions on the mailer, instead of sending them entirely anonymously and this is the only reason we are even hearing about this study. Had they sent it anonymously as most other scholars have there might have been some complaints in Montana, but no traceable path back to Stanford or Dartmouth.

Much of the debate about this study involves the use of the state seal and the partial lack of IRB approval. Let me first note, that compared with the types of electoral violations common in comparative politics - vote-buying, ballot stuffing, electoral violence and coercion these do not seem like the crimes of the century. Of course, scholars should not be running illegal experiments without IRB approval - but these might have just been oversights.

The experiment, however, raises two bigger issues here that apply to all experiments like these: one is the potential that the experiment might

impact a real election, and the other is the lack of informed consent. These are common in all electoral field experiments as well as many other types of field experiments. These questions are often ignored by the field, but are now being discussed.

Lets consider first the question of conducting a field experiment during an actual election. Most political scientists I've talked to cringe at the thought of affecting a real election. We have great respect for and hold sacred the institutions of democratic elections. For many of us, there is something uncomfortable about affecting the outcome. Others strongly disagree.

There are several arguments in favor of research that has an impact in the real world, including in elections. One is that everything we do has an impact. An observational study might have a major policy impact and change the way governments or politicians behave; why should we worry about changing voters' behavior? A goal of empirical science is in fact do just that impact the real world. Regarding elections, many other political actors are doing exactly the same things as political scientists - mobilizing and persuading voters - but with less altruistic motives than our own. Perhaps we should not concerned about treatments that are low-risk and common in the real world.

Further, one can argue - though there is a counter argument - that our treatments are normatively good - that we contribute to democracy directly through our experiments. If a treatment provides accurate factual information, then our experiment has the benefit of increasing voter information. How can this be anything but a good thing? If our treatment is mobilizing voters, increasing participation of underrepresented groups, many would defend this as a valuable contribution to democracy.<sup>1</sup>

Others are opposed to such interventions. One reason is that manipula-

---

<sup>1</sup>Of course, political information is highly contextual. For example, one election study told voters that a candidate had been accused of a crime, but never told them that they had not been convicted of anything. In the Montana Study, the mailer showed estimated judicial candidate ideal points, based on who had donated to them, using Bonica's method. But no doubt these were estimated with some uncertainty. Indeed, in some cases the two candidate might have been statistically indistinguishable from each other, though they were presented without any uncertainty. Why not provide confidence intervals? The answer, of course, is that it would overly complicate a simple experiment. In other words the PIs may have actually provided inaccurate information in an experiment. The point estimates may have been correct, but reported without context makes the information misleading and incomplete. Even simple factual information - statistics calculated using a proven algorithm - can be misleading.

tions during elections always result in harm to someone, because elections are almost always zero-sum games (Zimmerman, 2013)<sup>2</sup>. Consequently, any impact of an experiment on election results inevitably benefits one candidate and harms another. At the extreme, we might take an ethnographers position that we must leave no trace, as Dawn Teele points out (?), a version of the Prime Directive?.

A second question is regarding the normative value of these treatments. I doubt many scholars would complain if a low-risk treatment - like a simple mailer - reduced corruption, increased literacy, lowered cholesterol, or reduced secondary hospital infections. There is a general agreement that these are normatively valuable for human society; perhaps they are valence goods in the language of some of our theorists.

But what is the normative value of persuading voters to support one candidate or another? Even if one accepts the benefits of increased information (a treatment), the normative value of the dependent variable (vote choice) is less clear. Reasonable people can support candidates on either side of most political spectrums, and a central part of democracy is tolerance for diversity of opinion. This means that changing voting behavior cannot be defended as delivering benefit to subjects. It simultaneously harms and benefits because, as Zimmerman (2013) points out, politics is a zero-sum game<sup>3</sup>. An increased vote share for one candidate means fewer votes for someone else. The degree of cost associated with a change in vote share depends on many features of the context: electoral rules that turn seats into votes, the competitiveness of the race and of the legislative body (if a legislative race), whether campaign funding or other resources are distributed in future campaigns according to vote share, for example.

Are turnout field experiments (GOTV experiments) are normatively valuable if they increase participation in politics? These can be criticized as well. Since we know that voting is irrational and that any individuals vote is very unlikely to affect an outcome, GOTV experiments are effectively encouraging people to waste their time instead of spending it with family, friends, and fun. Suppose, for example, that it takes 30 minutes to vote, and our

---

<sup>2</sup>In some mixed member systems, the number of legislators is partly determined by the election results. So when just measuring candidate harm and benefit, the tradeoffs are not always zero sum.

<sup>3</sup>One might argue that in some contexts, we can make normative statements about one candidate or another, perhaps in the German elections of 1930 and 1932. But in most consolidated democracies, such judgements are not appropriate.

treatment mobilizes 10,000 subjects. Did we just waste 7 months of human time? Possibly.

This means that when we mobilize subjects to vote with GOTV messages, or try to persuade them to vote for one candidate or another, we are intervening in their lives, possibly affecting the lives of all individuals in the polity defined by the electoral district, and the normative benefits of doing so are quite ambiguous.

The question of affecting elections has a practical side as well as an ethical one. Many citizens and politicians don't like the idea of scientists experimenting during real elections. As a discipline of largely public servants supported by public money, it seems ill-advised to make our principals angry. It's easy to imagine that the level of anger might be correlated with the size of the experiment, the origin of the research funds, and the impact on the election. A massive study funded by a foreign government that changes an election will generate more outcry than a small sample with domestic funding.

For most people, all of the above problems are conditional on a more fundamental question: must we tell our subjects that we are experimenting on them? Frequently in field experiments, subjects and bystanders often do not even know that they are in a study. This violates the principle of informed consent, which means that all subjects must be told about the nature, aims, and risks of the study, and only after being informed, they must give voluntary consent. Historically, most of the great scandals of human research involve a lack of informed consent, including Milgram's Obedience to Authority and the Tuskegee Syphilis study. For this reason, informed consent is required in almost every research study with human subjects. Scholarly experiments can only skip informed consent when the risks to subjects are minimal and the research could not be carried out otherwise (45CFR46.116.d).

Let me also emphasize that intervening in elections requires thinking not only about subject, but also thinking about bystanders. Foos and de Rooij (2014) shows how GOTV interventions spread through households so more than just those contacted are affected. But everyone in the jurisdiction where the treatment is administered is potentially an affected bystander if the election result is changed, then we've affected democracy in that polity, for better or for worse.

How are these questions related in field experiments? Whether or not impact is an issue depends on whether or not subjects are informed and consenting. Suppose, hypothetically, that everyone in Montana had been invited to participate in a study where they would receive some mailers, and

suppose that they only received a flyer if they had agreed to participate. I doubt anyone would be upset if all the subjects knew they were in a study and all had consented to receive a treatment.

This suggests that the most critical issue to resolve is not whether we are impacting the real world, but whether our subjects are aware that they are subjects. If we can resolve the consent question, we can worry less about the impact question. Unfortunately, fully informed consent might render some hypotheses difficult to test.

One possible solution is exploring other types of acceptable consent, as discussed by Humphreys (2014) . He proposes a series of alternative forms of consent, for example, superset consent: listing many possible treatments and having the subject consent to the entire set, not knowing which one of them will be used in the experiment. Another option is proposed by bioethicist Koenig (2014) who has argued that in some cases, consent could be granted by a representative citizens body. A third possibility is part of federal human subjects regulations but is rarely used by political scientists: instead of seeking consent, simply inform all subjects after the study. Presumably scholars would anticipate subjects reactions and avoid controversy in experimental design.

Of course, there are two other alternatives as well: do it in a lab or in a survey experiment, or at least thoroughly explore the effect of the treatment in these environments before going to the field. Most of these studies can be tested in laboratories or in survey experiments. They will lose an element of realism, but gain informed consent. After several replications, then consider whether a field experiment will provide anything more than cachet.. Field experiments are not without critics for their scientific benefits, though they are currently in vogue. We can still learn form lab and survey experiments, though they do not always get as much attention as an abstract that includes a statement that we treated 100,000 subjects in a real election.

Some of these alternatives might be more efficient designs - more power, lower cost, and an ability to test more hypotheses. Suppose the cost of sending political mail to 100,000 subjects is \$100,000. That amount should easily supported a survey experiment with several thousand subjects and with post-election vote validation. The survey experiment would have had more covariates and no controversy. It might have produced a larger final sample size since the unit of analysis in the field experiments is often the precinct or electoral district, not the individual.

There are clearly alternatives to conducting field experiments, and al-

ternatives may be less expensive and more powerful. They do not have all the benefits of field experiments - they lack realism and can suffer from Hawthorne Effects. But they can allow scholars to avoid controversy while getting the research done.

However, if one is committed to conducting a field experiment in an election, I suggest several standards for how we should do this kind of research. If you are going to intervene in a real election, without informing subjects, consider adopting these design features to assuage ethical and practical concerns:

1. Minimize the risk that the intervention will affect the election. Pick elections where the challenger doesn't have a chance. If it gets close, get out. Pay attention to long-term consequences of a study. For example, future campaign funding may be a function of past elections vote share so any intervention harms one party over another.
2. Do a power analysis and minimize the size of the subject pool. There's no reason to have 100,000 subjects when 5,000 would be enough for most studies.
3. Compensate subjects, directly or indirectly. Directly, one could approach and pay all subjects after a study. In the Montana Study, this might mean mailing \$2 to all subjects after the election to compensate them for their time. Alternatively, it might mean paying \$2 per subject into some benevolence fund, perhaps a donation to a school or other institution in the treated communities. This shows respect for subjects' time, and it also serves as a constraint on scholars who might otherwise maximize the size of the study and its impact without any real need.
4. If your study involves deception or a lack of consent, inform all your subjects, after the study, that they were in a study. This is listed as a requirement when appropriate in the Common Rule, but I've never seen anyone do it. Consider as well reporting the magnitude of the experiment's impact on their election. This again serves as a constraint on scholars. Anticipating subjects' responses to the study, they may preemptively move to avoid controversy. It is one thing to run a study without consent for scientific reasons, but avoiding publicity after the study is over suggests that we have something to hide, and that we know internally that we have acted inappropriately.

5. If you are going to treat voters with information during a real election, be sure to provide factual, balanced, and informative messages, not half-truths or other misleading information.
6. Dont break the law. I suspect that few scholars go into the field intending to commit electoral crimes. But we arent election lawyers and political communication is highly regulated in many systems. This implies significant due diligence by PIs before going into the field.
7. Humphreys (2013) suggests that careful partnerships with other political actors that normally do intervene in elections can help us avoid damage to the discipline and inappropriate interventions. This is a nice intermediate suggestion between restricting ourselves to laboratory or survey experiments, and field experiments. It does require careful assessment of the scholar's relationship with the third party, as discussed by Hyde and Nickerson (2014).

I will conclude with four recommendations for moving forward.

First, as a discipline we should engage these issues directly and work toward shared norms. The past several days have shown that political scientists have widely differing views on these issues. Yet many scholars do not want to even talk about ethics. This is unfortunate. The future of field experiments as a viable way of testing hypotheses about politics and political behavior depends on engaging directly with these issues. This is especially important because there are many other ethical issues emerging across experimental political science<sup>4</sup>. We need more dialogue and research on these issues to avoid harm to subjects and to avoid damaging the reputation and future of the field.

Second, as argued by Zechmeister (2013), we cant outsource ethical judgments to IRBs. IRB approval is neither a blank check nor absolution from sins past, present, and future. IRB's may do their best (or worst), but scholars are the experts who know their cases and must take responsibility for keeping their research legal and ethical.

Third, we can make small institutional and norm changes that will help us avoid scandal, while avoiding more regulation. One would be posting all our IRB applications online when our work is published. Another would be

---

<sup>4</sup>See, for example, Nielsen (2014); Dionne, Harawa and Honde (2014); Merolla (2014); Lü (2014)



adopting norms that, in field experiments without informed consent, subjects will be advised post-treatment that they were in a study and core results presented. A third change could be a norm of compensation for subjects, even if indirect. The alternative to making some changes is that at some point, after several more scandals, we are likely to face stricter and stricter IRB's.

Lastly, we all need to acknowledge that when it comes to ethics and research, each and every one of us has a conflict of interest. There is always a trade-off between protecting human subjects and executing our research agendas. Protecting subjects will almost always make research harder. There are strong career incentives to overlook rules and risks and push for bigger, more aggressive treatments that have a better chance of being published. Its easy to convince ourselves that the costs to subjects are low and the benefits to society are high even when they aren't.

## References

- Bonica, Adam. 2014. "Mapping the Ideological Marketplace." *AJPS* 58(2):367–387.
- Dionne, Kim Yi, Augustine Harawa and Hastings Honde. 2014. The Ethics of Exclusion when Experimenting in Impoverished Settings. In *Ethical Challenges in Political Science Experiments*, ed. Scott Desposato.
- Foos, Florian and Eline de Rooij. 2014. "Household Partisan Composition and Campaign Mobilization."
- Humphreys, Macartan. 2013. "Manipulation and Consent." Paper presented at the UCSD Conference on Political Science Experiments.
- Humphreys, Macartan. 2014. "How to make field experiments more ethical." Post to The Monkey Cage Blog.
- Hyde, Susan and David Nickerson. 2014. Conducting Research with NGOs: Relevant Counterfactuals from the Perspective of Subjects. In *Ethical Challenges in Political Science Experiments*, ed. Scott Desposato.
- Koenig, Barbara. 2014. "Have we asked too much of Consent?" *Hastings Cent Rep* 44(4):33–4.
- Lü, Xiaobo. 2014. China: Ethical Challenges in Comparative Politics Experiments. In *Ethical Challenges in Political Science Experiments*, ed. Scott Desposato.
- Merolla, Jennifer. 2014. The Value and Challenges of Using Local Ethical Review in Comparative Politics Experiments. In *Ethical Challenges in Political Science Experiments*, ed. Scott Desposato.
- Nielsen, Richard. 2014. Ethics for Experimental Manipulation of Religion. In *Ethical Challenges in Political Science Experiments*, ed. Scott Desposato.
- Zechmeister, Elizabeth. 2013. Ethics and Research in Political Science: The Responsibilities of the Researcher and the Profession. In *Ethical Challenges in Political Science Experiments*, ed. Scott Desposato.

Zimmerman, Brigitte. 2013. Ethical Concerns Surrounding Research Interventions in Democratic Processes. In *Ethical Challenges in Political Science Experiments*, ed. Scott Desposato.