The Ethics of Electoral Experimentation: Design-Based Recommendations

Tara Slough*

June 7, 2019

Preliminary. Please find the most recent version here.

Abstract

Experimental interventions on elections offer insights about the preferences and behavior of voters and politicians. Such experiments represent represent an increasingly popular tool used by academics, campaigns, and corporations (e.g., Facebook). However, by manipulating aspects of campaigns or elections, experimenters can also change who wins office. This represents a critical ethical concern in the use of electoral experiments. I argue that researchers should design experiments to minimize the number of election outcomes that their intervention could plausibly change. I derive an upper bound on the possible effect of an experimental intervention on vote share. Using these bounds, I formulate guidelines for the design of electoral experiments around this ethical objective.

*Ph.D. Candidate, Columbia University, tls2145@columbia.edu. This project is supported in part by an NSF Graduate Research Fellowship, DGE-11-44155.

1 Introduction

Experiments on real elections represent an increasingly common tool in studies of elections, political behavior, and political accountability. While the use of experiments on elections dates back nearly a century (Gosnell, 1926), the scale, sophistication, and frequency of elections experiments has increased precipitously since the late 1990s. A central ethical concern in the study of elections is that by manipulating characteristics of campaigns, candidates, or voter information, researchers may also be changing aggregate election outcomes.

Two notable changes since the pioneering experimental studies of elections by Gosnell (1926), Eldersveld (1956), Blydenburgh (1971), and Gerber and Green (1999, 2000) influence these ethical considerations. First, researchers now work largely *outside* jurisdictions where they work or plausibly reside. As a discipline, we now work far beyond elections in Chicago, Ann Arbor, and New Haven. In this sense, researchers will not typically internalize the costs or benefits of having altered an election outcome, should their intervention do so. However, the subjects of their experiments will internalize these outcomes.[1] Second, in addition to academic researchers, campaigns and technology companies now regularly implement massive experimental interventions in elections (see, for example, Pons, 2018; Bond et al., 2012).

I focus on the ethical concern that experimental manipulations may alter aggregate election outcomes. This concern is not new. For example, Dunning et al. (2019) write that the authors of six coordinated experiments on elections and accountability "elaborated research designs to ensure to the maximum extent possible that our studies would not affect aggregate election outcomes" (52). However, it remains conspicuously absent from disciplinary discussions of experimental ethics (e.g., Desposato, 2016). Indeed, this consideration appears to be invoked informally, if at all, in most *ex-post* accounts of electoral interventions. This article proposes a formal, design-based approach to the *ex-ante* consideration of how experimental interventions could affect aggregate election outcomes.

[1] This is a feature of all electoral interventions to be sure, yet we may be particularly attuned to such issues when the "distance" between experimenters and subjects increases.

The ethical considerations related to experimental research on elections are admittedly far more complex than the focus on aggregate electoral outcomes in this paper. Such a calculus requires consideration of the benefits reaped from such research – from knowledge to the development of reforms/tools designed to promote political accountability – in addition to practices intended to protect subjects and the implementers. While these points merit a lengthier discussion, the focus of this paper is to consider how experimenters can minimize the risk of changing election outcomes.

This paper identifies two features of interventions that should inform ethical considerations of electoral interventions. First, under standard assumptions invoked in the analysis of experimental data, the *saturation* of an intervention within an electorate should determine the degree to which we may plausibly expect an experimental treatment to move electoral returns in the population. Second, levels of an electoral outcome, i.e. turnout, in the absence of treatment place bounds on the maximum degree to an intervention could change voter behavior (Manski, 2003). In tandem, these features allow researchers to estimate, *ex-ante*, bounds on the degree to which an intervention could change outcomes. I propose that these bounds be compared to the distribution of electoral outcomes in previous elections to estimate the maximum share of aggregate outcomes an intervention could plausibly change as a function of a given intervention.

One implication of this analysis is that the ethics of experimentation with respect to aggregate election relates closely to electoral institutions and competition. In general, an experiment in a small electorate is more likely to alter aggregate outcomes than an otherwise equivalent experiment in a larger electorate. Further, experiments in more competitive races are riskier than experiments in landslide races. To the extent that these considerations weigh into experimental design, they also posit implications for the external validity of experiments designed to reduce these ethical concernsb.

I further identify a tradeoff between these ethical considerations and statistical power in the design of such experiments. Voter-level outcome data allows researchers to target interventions to the individual or household, allowing for a sparser distribution of treatment. However, when outcome data is aggregated to the electoral table, precinct, or municipality level, researchers typi-

cally respond with clustered designs. Such clustered designs typically need to be administered to more voters (individuals) to achieve adequate power. Yet, this implies a much denser allocation of treatment, raising concerns about the degree to which such interventions may change electoral results.

The tools developed here serve to provide guidance for researchers considering prospective interventions on elections as well as consumers of research built upon such interventions. It advances the view that ethical considerations should be more nuanced question than whether an intervention should be done. Instead, it should be a question of how an intervention could be designed to minimize the threat of harm to subjects – here, a function of who wins political office.

2 The Ethical Objective

I focus on ethical considerations unique to experiments on elections, abstracting from broader ethical considerations for human subjects research or experiments. I assume that researchers' objective is to avoid changing who ultimately wins office through an experimental manipulation. In the aggregate, thus, researchers would ideally minimize the probability that their interventions change the *ex-post* distribution of seats or offices.[2] In so doing, I assume that the primary electoral consequences on policymaking or governance occur because candidate A wins office, not because candidate A won office with 60 percent instead of 51 percent of the vote (no mandate effects).

The approach advocated here considers two types of uncertainty that we have as researchers. First, we lack the omniscence to determine whether electoral outcome is normatively better than another for constituents or the grounds upon which such a determination could be made. Indeed any electoral outcome is apt to produce winners and losers. The approach here simply asserts that researchers should not be determining who wins and who loses in the service of research. Second, we do not know what an election outcome would be in the absence of an experimental manipulation. This limits our ability to design an experiment to minimize the probability that their

[2]In principle, this consideration is be also relevant to interventions on legislators who craft policies (e.g., Zelizer, 2018; Malesky et al., 2019). The outcome of elections, however can be seen as plausibly encompassing all debates, sponsorship, or roll call voting etc. in a legislative session, not simply a "treated" piece of legislation.

interventions change the *ex-post* distribution of seats or offices. As such, this paper advocates the estimation of conservative bounds on the *ex-ante* possible shift in vote share. These bounds can be calculated analytically and compared to distributions characterizing relevant measures of closeness in elections.

When does an elections experiment become unacceptable on grounds that it is too likely to change election outcomes? In principle, we could eliminate the risk of influencing electoral outcomes entirely by not running these experiments (Beerbohm, Davis, and Kern, 2017). Yet, we also learn about political behavior, persuasion, and electoral accountability from these interventions. Some existing experimental interventions are small (or sparse) enough to have a near-negligible effect on electoral outcomes, even by the convservative standards specified in this article. Accepting that there are some benefits to this research, this paper does not specify a threshold over which experiments become "unethical." Instead it provides a more systematic way to bound possible effects *ex-ante* such that research can be designed (or avoided) in a way to greatly reduce these risks. By reporting these quantities in grant applications, pre-analysis plans, and ultimately research outputs, researchers can transparently justify the their design choices.

3 Design-Based Considerations

3.1 Outcomes: Level of Aggregation

Experimental interventions on elections typically study the effects of an electoral intervention on turnout and/or some measure of vote choice. One measure of outcomes consists of self reports of turnout and vote choice measured using endline surveys. However, self reports are known to overstate rates of turnout by substantial margins (Burden, 2000; Jackman and Spahn, 2019), similar distortions relative to behavioral measures are likely for vote choice (Boas, Hidalgo, and Melo, 2019). In interventions encouraging turnout or voter sanction of poor-performing politicians, we might expect levels of misreporting to be correlated with treatment assignment as a result of demand effects.

Given the limitations of survey data, we increasingly rely on administrative electoral outcome

data. However, these results are typically aggregated to (at least) the precinct level. One notable exception is the measure of turnout using the US Voter Data files. Unfortunately, individual level administrative turnout data is not available in many other contexts. Where such data is unavailable and for the analysis of vote share outcomes, we measure outcomes at an aggregate level. In general, the move to aggregate outcomes induces researchers to use cluster-randomized experiments where the cluster is the unit of administrative data aggregation (precinct or jurisdiction). The move from individual- to cluster-randomized designs often requires administration of treatment to more voters in order to detect effects.

3.2 Treatment Assignment Density

This paper focuses posits a central role of treatment assignment density, or saturation, within an electoral district. This quantity captures the proportion of registered voters in a district that are assigned to (a) treatment. Importantly, this measure of treatment assignment density is not a standard consideration in the design of experimental interventions. In individually-randomized electoral experiments, the proportion of subjects assigned to treatment is, in general, a small fraction of all voters in a district.[3] In cluster-randomized electoral experiments, clusters are typically coterminous with or smaller than electoral districts. For example, interventions aimed to change incentives for candidate entry imply that clusters and districts must be coterminous (e.g Gulzar and Khan, 2018). In contrast, in experiments that assign information (e.g. flyers) to precincts, voters in a cluster generally represent a proper subset of voters in an electoral district.

One point to which I will return in Section 5.1 is that it may be the case that in an election, voters vote upon candidates or ballot issues in distinct districts on the same day as is common in US elections. In this case, the treatment assignment density may vary across races on the same ballot. To the extent that interventions could affect turnout, it may be important to determine the highest level of saturation of any race.

[3]If every registered voter in a district was assigned to treatment or control (no placebo), the treatment assignment density would be the proportion of registered voters assigned to treatment.

4 Bounding Effects on Electoral Behavior

In deriving bounds for the effect of an experimental intervention on electoral outcomes, I distinguish aspects of the design that are "controlled" by the researcher from voter responses to a treatment which are beyond the researcher's control. In deriving these bounds, I invoke two assumptions that are generally invoked in the analysis of randomized experiments, as follows. First, treatment assignment is ignorable,[4] which is ensured by the random assignment of treatment. Second, I invoke the stable unit treatment value assumption (SUTVA), which implies non-interference between units and a single version of each treatment.[5]

4.1 Defining the Saturation of the Experiment

I proceed by defining the saturation of the experiment relative to the electorate in a given election. Here, I consider the mapping of the unit at which treatment is assigned, indexed by $j \in J$, and the electoral district, $d \in D$. In an experiment that assigns individual voters to some treatment, j corresponds to individuals; in a cluster-randomized experiment, j corresponds to a cluster. Denote the number of registered voters in j as $n_j \geq 1$ and the number of registered voters in a district d as n_d, where $n_d = \sum_{j \in d} n_j$.

The treatment may be administered to any subset of units in a district, $S \subseteq D$. Define the saturation of an experiment in a district, $\mathcal{S}_d$, as the proportion of voters sampled (whether as an individual or a member of a cluster) for the experiment. Formally, this quantity is:

$$\mathcal{S}_d = \frac{1}{n_d} \sum_{j \in S} n_j \tag{1}$$

Given SUTVA and a pure control condition, the degree to which treatment may change electoral results should depend on the proportion of subjects assigned to treatment. Denote the proportion of voters in unit j that are assigned to treatment as π_j.[6] In most individual and cluster-randomized

[4] Formally, $Y(z) \perp Z$ and $Pr(Z = z) > 0 \forall z$.

[5] Formally, $Y_i(z_i) = Y_i(z_i, \mathbf{z}_{\neg i})$.

[6] In an experiment with multiple treatment arms, π_j should represent the share of voters in unit j that are assigned to any non-pure control condition.

experiments, $\pi_j \in \{0, 1\}$. One notable exception is cluster-randomized multilevel or saturation designs where the goal is to treat some proportion of individuals in a cluster such that $\pi_j \in (0, 1)$ (Sinclair, McConnell, and Green, 2012). The treatment saturation, $\mathcal{S}_d^T$ can then be defined:

$$\mathcal{S}_d^T = \frac{1}{n_d} \sum_{j \in S} \pi_j n_j \tag{2}$$

4.2 Bounding Plausible Treatment Effects

I proceed by bounding the range of plausible treatment effects. At the individual (voter) level, turnout and vote share outcomes can be categorized as a binary decision. For example, turnout measures a decision to vote or not vote; vote share measures a decision to vote for some option/candidate A as opposed to any other outcome, $\neg A$. Given this classification, the average treatment effect (ATE) is the difference in the proportion of citizens that vote in treatment versus control. This implies ATE $\in [-1, 1]$ for any level of aggregation.

However, researchers often have access to baseline data on the outcome of interest, e.g., turnout in election $t - 1$. If turnout in election $t - 1$ provides an accurate estimate of turnout in the control group in election t, then a tighter bound can be derived. Denote the approximation of the expectation of the control potential outcome as $E[Y_i(0)]$, where i indexes individual voters. In the case that outcomes are measured at the level of unit j, one can bound ATE$_j \in [-E[Y_{ij}(0)], 1 - E[Y_{ij}(0)]]$. Note that these bounds are effectively "extreme value" bounds (Manski, 2003).

Note that an ATE of 0 implies no aggregate effect on election outcomes. One implication of the bounds described here is that if $E[Y_{ij}(0)] \neq \frac{1}{2}$, the bounds are not symmetric about 0. For the analysis that follows, denote the bound (lower or upper) with the greatest distance from 0 as τ_j:

$$\tau_j = \max\{E[Y_{ij}(0)], 1 - E[Y_{ij}(0)]\} \tag{3}$$

$$= \begin{cases} E[Y_{ij}(0)] & \text{if } E[Y_{ij}(0)] \leq \frac{1}{2} \\ 1 - E[Y_{ij}(0)] & \text{if } E[Y_{ij}(0)] > \frac{1}{2} \end{cases} \tag{4}$$

Note that I am agnostic about the direction of treatment effects. Indeed, in experiments on retrospective accountability as in Dunning et al. (2019) and other works, unless the information (the signal) falls outside of the support of the distribution of voter priors, models of Bayesian updating provide no theoretical expectation of monotonic treatment effects. The bounds chraracterized in Equation 3 make no assumption about direction of treatment effects.

Taken together, I define the maximum aggregate electoral impact, by district as:

Definition 1. *Maximal Aggregate Electoral Impact: The ex-ante maximal aggregate electoral impact (MAEI) in district d is given by:*

$$MAEI_d = \frac{\sum_{j \in S} \pi_j n_j \tau_j}{n_d} \tag{5}$$

Comment 1. *Comparative statics. The $MAEI_d$ is decreasing in the size of the electorate, n_d; increasing in the saturation of treatment, $\pi_j n_j$; and increasing in the bound on changes in vote share τ_j.*

The $MAEI_d$ is straightforward to calculate given data on the size and structure of the electorate. Where data is insufficient to approximate $E[Y_{ij}(0)]$ in cluster-randomized experiments, one may assume conservatively that $\tau_j = 1$.[7] Comment 1 posits several immediate implications. Most obviously, an identically designed experiment has less possibiity of moving aggregate vote share or turnout in a large district than in a small district. In other words, the bounds we can place on the electoral impact of the same experimental design are much narrower for a presidential election than for a local school board election. However, due to researchers' desire to work in low-information contexts, much recent focus has been placed on legislative or local elections. This result suggests that this decision carries greater risks of altering turnout or vote share.

Second, Comment 1 suggests that higher saturation of treatment implies greater potential effects on vote share. This introduces a tradeoff between statistical power and the degree to which an experiment could alter aggregate electoral outcomes. Treating more individuals increases the

[7]Note that in individually-randomized experiments, $\tau_j = 1$.

saturation of treatment, possibly moving more votes. Moreover, a move from a individually-randomized to a cluster-randomized experiment requires many clusters for adequate power to detect effects. To the extent that researchers treat all individuals in clusters, the saturation of treatment increases substantially. Thus, one implication of lack of individual-level outcome data is that the possibility of electoral interventions increases substantially.

5 Assessing the Consequences of Electoral Interventions

The final implication of Proposition 1 with respect to τ_j demands a discussion of the ability of electoral experiments to change electoral outcomes, that is, who wins. While analyses of electoral experiments typically focus on vote share, not probability of victory (or seats won in a PR system), the lever through which elections have consequences is who wins office.[8]

The mapping of votes to an office or (discrete) seats implies the existance of at least one threshold, which, if crossed yields a different realization of officeholding. For example, in a two candidate race without abstention, for example, there exists a threshold at 50 percent. It is useful to denote the "margin to pivotality," ψ_d, as minimum change in vote share at which a different officeholder would be elected in district d. In a plurality election for a single seat, this is the margin of victory. In a PR system, there are various interpretations of ψ_d. Perhaps the most natural interpretation is the smallest change in any party's vote share that would change the distribution of seats. If $\psi_d > MAEI_d$, then an experiment could not change the ultimate electoral outcome. In contrast, if $\psi_d < MAEI_d$, the experiment *could* affect the ultimate electoral outcome. Note that like $E[Y_{ij}(0)]$, ψ_d is not knowable in advance of an election. Yet, we may have more information at the district (or electorate) level than we have for smaller subsets of voters.

As such, researchers should seek to minimize the number of districts, D, for which $MAEI_d > \psi_d$. Ultimately, the researcher's objective can be formalized as $\min \sum_{d \in D} I[MAEI_d > \psi_d]$, where $I[\cdot]$ is an indicator function. This implies that researchers should try to reduce $MAEI_d$ and/or

[8]This analysis excludes so-called "mandate" effects, though if this were the objective, the conceptualization of the ethical optimization problem becomes more straightforward. In this case, researchers would simply want to minimize $\sum_d MAEI_d$, subject to relevant constraints.

increase ψ_d. Given the width of the Manski bounds on τ_j, the $MAEI_d$ can most "efficiently" be minimized by reducing the density of treatment within an electoral district. While we do not know ψ_d prior to an election, in some cases, we can make a more informed assessment than in others. In particular, if electoral outcomes in election $t-1$ are generally prognostic of electoral outcomes in election t, then researchers could compare the $MAEI_i$ to a previous $\psi_{d,t-1}$. Alternatively, in races with frequent, high quality public opinion polling, researchers could estimate ψ_d on the basis of pre-election (and pre-intervention) opinion polls.

In settings where past outcomes are not prognostic, there is much more uncertainty about how to estimate ψ_d when planning an experiment. In such cases, I propose comparing the MAEI to the distribution of outcomes in past (analogous) races. Of course, for this exercise, the covariance between ψ_d and the extreme value bounds becomes a critical consideration. Note that in two party (candidate) plurality elections where $d \equiv j$, $Cov(\psi_d, \tau_j) < 0$. In close races, when vote shares hover around 50%, the more distant Manski bound, τ_j is tighter than in distant races. The covariance structure is more ambiguous when $S \subset d$ or in races with ≥ 2 candidates.

5.1 Robustness to Assumptions

In general, the proposed approach to bounding the impact of an electoral intervention invokes very few assumptions and bounds should be conservative to the extent that observed *findings* of known electoral interventions are modest relative to τ. I argue that researchers should plan experiments around *conservative* bounds. In considering robustness to assumptions, I consider the situations in which these bounds could be insufficiently conservative. In other words, under what conditions is it possible that the estimate of the $MAEI_d$ is an underestimate?

If there exists interference between units, a violation of SUTVA, it is possible that an intervention could affect units beyond those in the treatment group. In the notation described earlier, a violation of SUTVA leads to an increase in the treatment saturation S_d^T. Notably, such spillovers could be modeled as an increase to the subset of experimental units, i.e. S, or the "portion" of a unit that is exposed to treatment, π_j. To the extent that spillovers may expand to non-experimental units, we could think of treatment as increasing n_j (if in the same district) or affecting voting

outcomes in other districts (introducing a new d). *Ex-ante* consideration of spillovers, thus, must increase the estimate of $MAEI_d$.

6 Implications for Research Design and Learning

The derivation of the $MAEI_d$ posits immediate implications for the design of electoral field experiments. First, the invocation of Manski (2003) bounds clarifies one core assumption: as researchers, we cannot control how subjects (voters) will respond to a treatment. Further, the width of these bounds suggests that the most effective ways to reduce the probability of influencing an election are relate to the choice of districts and the allocation of treatment within a district.

In researchers' selection of districts in which to implement designs, large electorates provide opportunities to treat a sufficient number of individuals or clusters to power an experiment while treating the votes of a negligible share of the electorate. While campaigns and social media companies have implemented experiments in national districts (i.e., US presidential elections), such experiments remain the exception, not the norm of most electoral interventions designed by academics and NGOs. This paper makes the recommendation of moving "up" the types of elections that we treat, with two caveats. First, voters are typically more informed about high level elections. To the extent that a wide literature seeks to examine the impact of information relevelation to voters, it may be the case that these topics are much harder to study in the context of national elections. Second, researchers must remain aware that by "treating" voters with respect to one race in concurrent elections, they may be influencing the results of other up- or down-ballot races.

Second, to the extent possible, researchers should implement interventions in "uncompetitive" races to the extent that they are identifiable *ex-ante*. This suggestion comes from increasing the "margin to pivotality," or ψ_d. In such races, the ability of an experiment to change who wins office is lower. However, the study of voter or politician behavior in landslide races or "core" districts, may be quite distinct from behavior in more competitive constituencies. In this sense, researchers face a tradeoff in terms of the possible risks to election outcomes and the generalizability of insights about behavior.[9]

[9]While critiques of the lack of external validity of experiments are widespread, the idea that

7 Conclusions

This note suggests that researchers, campaigns, and corporations should more carefully design electoral experiments in order to minimize the risk of changing electoral outcomes. Drawing upon a framework built upon nonparametric extreme value bounds, I suggest that researchers can best reduce ethical concerns through the choice of electoral district and allocation of treatment therein. I advocate against high saturation of treatment within districts. When electoral outcomes are observed in the aggregate, researchers turn to clustered experimental designs. However, when these clusters represent large shares of a district's electorate, the possibility for experiments to change elections is heightened. Moreover, experiments are inherently riskier in elections that would be close in the absence of treatment.

I advocate the incorporation of ethical considerations as a much more prominent guide to research design than is presently described. Ultimately, researchers should not be the arbiters of what makes a "good outcome" and, increasingly, they do not internalize the consequences of the elections they manipulate. Careful research design can allow researchers to continue to draw some insights from the experimental study of elections while providing more protections to the communities that they study.

ethical considerations may lead to a less "representative" sample is new, to my knowledge.

References

Beerbohm, Eric, Ryan Davis, and Adam Kern. 2017. "The Democrati Limits of Political Experiments." Working paper, available at https://scholar.harvard.edu/files/beerbohm/files/democratic_limits_of_political_experiments_eb_rd_ak.pdf.

Blydenburgh, John C. 1971. "A Controlled Experiment to Measure the Effects of Personal Contact Campaigning." *Midwest Journal of Political Science* 15 (2): 365–381.

Boas, Taylor, F. Daniel Hidalgo, and Marcus André Melo. 2019. "Norms versus Action: Why Voters Fail to Sanction Malfeasance in Brazil." *American Journal of Political Science* forthcoming.

Bond, Robert M., Christopher J. Fariss, Jason J. Jones, Adam D. I. Kramer, Cameron Marlow, Jaime E. Settle, and James H. Fowler. 2012. "A 61-Million-Person Experiment in Social Influence and Political Mobilization." *Nature* 489: 295–298.

Burden, Barry. 2000. "Voter Turnout and the National Election Studies." *Political Analysis* 8 (4): 389–398.

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

Dunning, Thad, Guy Grossman, Macartan Humphreys, Susan D. Hyde, Craig McIntosh, and Gareth Nellis, eds. 2019. *Information, Accountability, and Cumulative Learning: Lessons from Metaketa I*. New York: Cambridge University Press.

Eldersveld, Samuel J. 1956. "Experimental Propaganda Techniques and Voting Behavior." *American Journal of Political Science* 50 (1): 154–165.

Gerber, Alan S., and Donald P. Green. 1999. "Does Canvassing Increase Voter Turnout? A Field Experiment." *Proceedings of the National Academy of Sciences* 96 (14): 10939–10942.

Gerber, Alan S., and Donald P. Green. 2000. "The Effects of Canvassing, Telephone Calls, and Direct Mail on Voter Turnout: A Field Experiment." *American Political Science Review* 94 (3): 653–663.

Gosnell, Harold F. 1926. "An Experiment in the Stimulation of Voting." *American Political Science Review* 20 (4): 869–874.

Gulzar, Saad, and Muhammad Yasir Khan. 2018. "Motivating Political Candidacy and Performance: Experimental Evidence from Pakistan." Working paper.

Jackman, Simon, and Bradley Spahn. 2019. "Why does the American National Election Survey Overestimate Voter Turnout?" *Political Analysis* 27 (2): 193–207.

Malesky, Edmund, Jason Douglas Todd, Anh Le, and Anh Tran. 2019. "Testing Legislator Responsiveness to Citizens and Firms in Single-Party Regimes: A Field Experiment in the Vietnamese National Assembly." Working paper.

Manski, Charles E. 2003. *Partial Identification of Probability Distributions*. New York: Springer.

Pons, Vincent. 2018. "Will a Five-Minute Discussion Change Your Mind? A Countrywide Experiment on Voter Choice in France." *American Economic Review* 108 (6): 1322–1363.

Sinclair, Betsy, Margaret McConnell, and Donald P. Green. 2012. "Detecting Spillover Effects: Design and Analysis of Multilevel Experiments." *American Journal of Political Science* 56: 1055–1069.

Zelizer, Adam. 2018. "How Responsive are Legislators to Policy Information? Evidence from a Field Experiment in a State Legislature." *Legislative Studies Quarterly* 43 (4): 595–618.