Detecting Spillover Effects: Design and Analysis of Multilevel Experiments

Betsy Sinclair   University of Chicago
Margaret McConnell   Harvard University
Donald P. Green   Columbia University

Interpersonal communication presents a methodological challenge and a research opportunity for researchers involved in field experiments. The challenge is that communication among subjects blurs the line between treatment and control conditions. When treatment effects are transmitted from subject to subject, the stable unit treatment value assumption (SUTVA) is violated, and comparison of treatment and control outcomes may provide a biased assessment of the treatment's causal influence. Social scientists are increasingly interested in the substantive phenomena that lead to SUTVA violations, such as communication in advance of an election. Experimental designs that gauge SUTVA violations provide useful insights into the extent and influence of interpersonal communication. This article illustrates the value of one such design, a multilevel experiment in which treatments are randomly assigned to individuals and varying proportions of their neighbors. After describing the theoretical and statistical underpinnings of this design, we apply it to a large-scale voter-mobilization experiment conducted in Chicago during a special election in 2009 using social-pressure mailings that highlight individual electoral participation. We find some evidence of within-household spillovers but no evidence of spillovers across households. We conclude by discussing how multilevel designs might be employed in other substantive domains, such as the study of deterrence and policy diffusion.

Many randomized field experiments take place within social settings, where individuals may influence each other's tastes, beliefs, and behaviors. Communication between participants, bandwagon effects, and other social psychological processes may violate what Rubin (1986) has termed the "stable unit treatment value assumption" (SUTVA), which is routinely invoked when drawing causal inferences about experimental effects. SUTVA holds that there is no interference between units; the experimental assignment of one subject has no effect on other subjects' potential outcomes.[1] SUTVA rules out "spillover effects" that occur, for example, when treated individuals transmit the information contained in the treatment to the control group (Rosenbaum 2007). Other examples of SUTVA violations outside the realm of elections include the displacement of crime from treatment areas that receive heightened police surveillance to control areas (Sherman and

---

[1] SUTVA holds that "the potential outcomes for any unit do not vary with the treatments assigned to any other units, and there are no different versions of the treatment" (Rubin 1986, 961). Our focus is on the first part of the SUTVA definition. Under this assumption, treatments are stable in the sense that potential outcomes are unaffected by the particular way in which subjects are allocated to experimental conditions.

Weisburd 1995), social comparisons that cause the control group assessments to be influenced by the intervention received by the treatment group (Sobel 2006), strategic interaction between subjects such that the control group adjusts its behavior in light of prior treatments and treatments received by others (Bednar et al. 2010), and strategic calculations that lead political actors in one jurisdiction to take cues from neighboring jurisdictions that receive a treatment, such as financial audits or election monitoring (Hyde 2010; Silva 2010).

Although SUTVA is fundamental to causal analysis, experiments (as well as observational studies) have typically downplayed the possibility of spillovers. In recent years, however, experimental researchers in political science (Nickerson 2008), economics (Miguel and Kremer 2004), sociology (Hong and Raudenbush 2006), and criminology (Ratcliffe et al. 2011) have devoted increasing attention to the substantive phenomena that lead to SUTVA violations and the statistical implications of these violations. Experimental designs that have the capacity to detect SUTVA violations therefore provide both substantive and methodological insights.

This article illustrates the value of one such design, a multilevel experiment in which treatments are randomly assigned to individuals and varying proportions of their neighbors. We begin by discussing SUTVA and the empirical literature that addresses it. After discussing the theoretical and statistical underpinnings of a multilevel experimental design, we apply it to a large-scale voter-mobilization experiment conducted in Chicago during a special election in 2009 using the social-pressure mailings pioneered by Gerber, Green, and Larimer (2008). Although our experiment has ample power, we find little evidence of cross-household spillovers, suggesting that interpersonal communication among neighbors was relatively limited. We conclude by discussing how this design might be employed in other domains, such as the study of deterrence and policy diffusion.

SUTVA and Related Literature

The Rubin Causal Model (Rubin 1974, 1978, 1980) defines the effect of a treatment on individual i as the difference between two potential outcomes: the outcome Y_{1i} that manifests when this individual receives the treatment ($t = 1$) and the corresponding outcome Y_{0i} when the treatment is withheld ($t = 0$). Randomized experiments are often motivated and interpreted using this framework. When individuals are randomly assigned to experimental groups and when the potential outcomes for a given individual are unaffected by the experimental assignments

of other individuals, the difference in average outcomes between the treatment and control groups provides an unbiased estimate of the true average treatment effect.[2]

The unbiasedness of simple difference-in-means estimation breaks down when potential outcomes for experimental subjects are influenced by the treatments assigned to others. Rather than characterize an individual's potential outcome (Y_{0i} or Y_{1i}) solely in terms of the experimental treatment that he or she receives, one must consider all of the potential outcomes that could occur when different combinations of subjects are treated. A more complete notation would allow for the consequences for subject i of treatments $t_1, t_2, \ldots, t_{i-1}, t_{i+1}, \ldots, t_n$, where subscripts index subjects 1 to n and where t_i refers to the treatment administered directly to the subject i. For any reasonably large experiment, the number of potential outcomes for each subject becomes extremely large.

In order to appreciate the potential for bias when SUTVA violations occur, consider the case of an experimental vaccination campaign (cf. Haber 1999; Halloran and Struchiner 1991, 1995; Hudgens and Halloran 2006; Miguel and Kremer 2004). Subject 1, a person who receives no vaccination but who is surrounded by other subjects who are vaccinated, is likely to have a different outcome than Subject 2, someone who is unvaccinated and is surrounded by other unvaccinated subjects. Both subjects are nominally part of the same experimental group, but the allocation of other subjects affects their potential outcomes. In an extreme case in which virtually everyone in the experiment receives the vaccine, the few stray unvaccinated subjects may avoid exposure to disease, causing the researcher who compares average outcomes in the treatment and control groups to underestimate the effectiveness of the treatment.

The challenge is to develop experimental designs that enable the researcher to measure spillover effects and gauge the effect of the treatment net of spillover effects. The objective is not solely to find significant spillover effects; so long as an experimental design has the capacity to detect spillovers, estimates of the treatment's direct effect are protected from bias regardless of whether spillover effects materialize.[3] If the experiment should reveal no

---

[2] Because our focus is on SUTVA, we have skipped over other assumptions and complications arising from attrition, noncompliance, and backdoor paths by which treatment assignment might affect outcomes other than via the treatment itself. For discussion of these issues, see Imbens, Angrist, and Rubin (1996).

[3] Other experimental designs that separate treatment and control groups geographically or temporally allow the researcher to ignore interference between units (David and Kempton 1996), but one advantage of our experimental approach is that we are able to estimate both direct treatment effects and spillover effects.

evidence of spillovers, the result suggests that researchers in this domain have less need to employ complex experimental designs in the future. Moreover, a null result sheds light on the absence of social dynamics that might otherwise be thought to generate spillovers.

Building on the recent literature discussing the identification of causal effects in the presence of interference between units (Hudgens and Halloran 2008), our approach is to gauge spillover effects using a multilevel experimental design. Multilevel experiments randomly vary whether a given subject receives the treatment and whether the treatment is administered to other subjects who are in some way connected to the subject. For example, in the study presented below, individual voters may receive a treatment; their housemates may receive a treatment; and others in their nine-digit zip code may receive a treatment. Persons, households, and zip codes therefore comprise the three "levels" in our experiment.

It should be stressed at the outset that multilevel designs do not "solve" the problem of satisfying the stable unit treatment value assumption. Given the vast number of ways in which treatments may be passed indirectly between subjects, multilevel design addresses certain facets of the SUTVA issue but cannot rule out all potential SUTVA violations. In the above example, two subjects in different households and different zip codes may nevertheless influence each other because they work for the same firm or attend the same church. Researchers confront an inherent identification problem when studying social influence. In order to make the empirical requirements manageable, researchers must impose some assumptions about the networks through which social influence is transmitted. One empirical strategy is to use multilevel designs to study what are thought to be the likely networks; if spillover effects are shown to be weak within these networks, the implication is that other channels that are not easily studied also have weak spillover effects. Of course, this is an inference, not a result, and more than one experiment may be necessary to determine which social networks, if any, transmit spillover effects.

To date, relatively few experimental studies have been designed with the explicit purpose of identifying spillover effects. One of the rare exceptions is the voter-mobilization experiment conducted by Nickerson (2008), which shows that the effects of door-to-door canvassing are transmitted within two-voter households. Nickerson finds that when one voter opens the door to a canvasser, 60% of the effect from the get-out-the-vote appeal is transmitted to the other household member.[4]

Apart from Nickerson (2008), the study of social influence is dominated by observational studies. Huckfeldt (1979) and Knoke (1990) use survey data to argue that individuals with politicized social networks are more likely to engage politically. McClurg (2003) suggests a mechanism, namely, information provided by others in the social network (see also Bolton 1972; Briet, Klandermans, and Kroon 1987; Gerlach and Hine 1970; McAdam and Paulsen 1993; McAdam 1986).

Researchers using hierarchical models and contextual data (Cho, Gimpel, and Dyck 2006; Cohen and Dawson 1993; Gimpel, Dyck, and Shaw 2004) have repeatedly demonstrated that political action is strongly predicted by neighborhood context and the behavior of others nearby. Although the claims generated by observational studies are plausible, this mode of analysis is potentially biased by what Manski (1993) calls the reflection problem: one's behavior may resemble that of one's social network because of shared unobserved influences that have nothing to do with social communication or any form of interpersonal influence. Studies such as Nickerson (2008) signal a movement away from observational approaches toward research designs that use randomization in order to identify causal parameters. Our modeling approach and empirical application elaborates Nickerson's work by gauging voter-mobilization spillovers both within and across households.

Model

As noted in the previous section, SUTVA violations are empirically intractable unless one imposes some theoretical structure on the network under study (Sobel 2006). This section presents a model that simplifies the identification problem and highlights certain problems of estimation. Anticipating the empirical application that follows, the model is developed in relation to a voter-mobilization effort in which mailings are sent to registered voters shortly before an election.

Individuals in this model are part of a social network. We focus on two specific network relationships that may be relevant for a voter-mobilization effort: the relationship between household members and the relationship between neighborhood residents. Each individual

---

[4] Other scholars have examined spillover effects in contexts unrelated to political behavior. Duflo and Saez (2003) use a multi-level design to study spillovers in individuals' choice of retirement plan; Sobel (2006) has examined spillovers in the Moving-to-Opportunity program; Hong and Raudenbush (2006) use a multilevel observational design to look at kindergarten retention; Hudgens and Halloran (2008) use a multilevel design to look at housing vouchers and vaccines; Miguel and Kremer (2004) study spillovers in the treatment of worms.

i belongs to a household h and a neighborhood z. For simplicity, we assume that individuals have the potential to interact with every member of their household and every member of their neighborhood but not with people outside their neighborhood. When one person receives a message from a mobilization campaign, the message may indirectly affect the voting probability among others in the household or neighborhood.

This indirect influence may be transmitted in a number of ways. People may discuss the campaign message. Or the message may cause some individuals to vote, and their behavior may signal the importance of the election to others in their network (Grosser and Schram 2006). Another possibility is that as people become more likely to vote, they publicize the election and lower the information costs incurred by others in their network. We remain agnostic about which of these mechanisms is at work and focus on their combined influence. We do, however, allow for the possibility that influence from household members has a different impact than influence from neighbors.

Identifying Direct Effects and Spillover Effects

In our model, the potential outcomes for each voter are a function of whether they are themselves treated, whether a member of their household is treated, and whether others in their neighborhood are treated. Anticipating the empirical application in the next section, we define direct treatment as 1 if a person is sent a mailing and 0 otherwise. We define household treatment as 1 if another person in the household is sent a mailing and 0 otherwise. And we define neighborhood-level treatment as 1 if all other households in the nine-digit zip code are sent a mailing, 0.5 if half of the nine-digit zip code's households are sent a mailing, and 0 if none of the nine-digit zip code's other households are sent a mailing.[5] We denote potential outcomes using the notation $y_{z,h,i}$, where z refers to the (nine-digit) zip code, h refers to the household, and i refers to the individual. Thus, y_{000} refers to a person whose zip code neighbors are untreated, whose housemates are untreated, and who receives no direct treatment.

In principle, for households with more than one voter, this notation allows for a total of $3 \times 2 \times 2 = 12$ potential outcomes for each person. In our empirical ap-

plication, however, we send mail to at most one randomly selected member of each household. This restriction means that we never observe y_{z11} outcomes, which leaves us with nine relevant potential outcomes. Another restriction is that in zip codes where all of the households receive mailings, there are no instances of y_{100}. Additionally, we vary the intensity of the neighborhood mailing to include a "low" intensity option, where just one other household in the zip code receives the treatment. This provides us with nine potential outcomes. These nine potential outcomes can be used to define several interesting estimands. If zip codes were to be untreated, the average direct effect of the mailings is $\bar{y}_{001} - \bar{y}_{000}$, while in half-treated zip codes it is $\bar{y}_{.501} - \bar{y}_{.500}$. If zip codes were to be untreated, the average household-spillover effect of the mailings on untreated individuals is $\bar{y}_{010} - \bar{y}_{000}$, while in half-treated zip codes it is $\bar{y}_{.510} - \bar{y}_{.500}$. Finally, the average zip-code-level effects can be defined as $\bar{y}_{.500} - \bar{y}_{000}$ or $\bar{y}_{.501} - \bar{y}_{001}$ or $\bar{y}_{101} - \bar{y}_{.501}$ or $\bar{y}_{110} - \bar{y}_{.510}$ or $\bar{y}_{.510} - \bar{y}_{010}$, depending on whether one is interested in the zip-code-level effects on the untreated, on the directly treated, or on those affected only by household spillovers, and depending on whether one wishes to estimate the effect of an increase in neighborhood saturation from 0 to 0.5 or from 0.5 to 1. Our particular multilevel experiment was designed to maximize the flexibility with which zip-code-level effects could be defined and estimated.

As noted earlier, we only observe one potential outcome for each individual; the other potential outcomes remain counterfactual. Random assignment of individuals to one of the experimental conditions, however, enables us to obtain unbiased estimates of the average treatment effects. For example, $E(\hat{\bar{y}}_{010} - \hat{\bar{y}}_{000})$ for the samples that are randomly assigned to these two conditions provides an unbiased estimate of the household-spillover effect on untreated individuals in untreated zip codes. A straightforward statistical test may be used to assess whether this average effect is distinguishable from zero, as well as whether the average household-spillover effect in untreated zip codes is distinguishable from the average household-spillover effect in partially treated zip codes. Multilevel assignment, in other words, enables us to identify direct treatment effects, household spillovers, neighborhood spillovers, and interactions among them. Moreover, this kind of experiment can be conducted using zip codes and households with varying numbers of voters, enabling the researcher to assess whether patterns of spillover differ systematically from one setting to the next.

This empirical strategy is more nuanced in what it can estimate and less prone to bias than experimental analyses that simply compare those who directly receive treatment

[5] In the experiment below, we further distinguish nine-digit zip codes where exactly one other household is treated. As intuition would suggest, the experimental results show no difference in voting rates between this experimental condition and the condition in which no other households in the nine-digit zip code are treated. Voting rates for each assignment category can be found in Table A1.

$(\bar{y}_{001}, \bar{y}_{101})$ with those who do not $(\bar{y}_{000}, \bar{y}_{010}, \bar{y}_{100}, \bar{y}_{110})$. Suppose, hypothetically, that each of these six groups were equal in size. The average voting rate in the treatment group minus the average voting rate in the control group would, in expectation, yield

$$\frac{1}{2}(\bar{y}_{001} + \bar{y}_{101}) - \frac{1}{4}(\bar{y}_{110} + \bar{y}_{100} + \bar{y}_{010} + \bar{y}_{000})$$

which does not correspond to the direct effect $\bar{y}_{001} - \bar{y}_{000}$ or $\bar{y}_{101} - \bar{y}_{100}$ or the average of the two. If the household-spillover effects are positive, the control group's voting rate will be elevated, leading to an underestimate of the average direct effect of treatment.

Although multilevel experimental design provides important insights into possible spillover effects, there are limitations in what an experiment of this kind can reveal about communication within social networks. Because we do not measure the communication flows within social networks, we cannot scale our estimated causal effects in terms of the influence per communication (Imbens, Angrist, and Rubin 1996). Even if we could measure communication, there remains the possibility that untreated individuals may read their housemate's mail, violating the identifying assumptions of statistical models designed to estimate the effect of communications among compliers (i.e., those who communicate only when encouraged by the experimental intervention). Nevertheless, a multilevel design does enable us to speak to important causal questions indirectly. If messages delivered to neighbors affect a subject's voting rates, we have strong circumstantial evidence for interpersonal transmission of information.

Statistical Power

Although multilevel designs broaden the range of hypotheses that can be tested, there is a downside: multilevel randomization reduces the statistical power with which the direct treatment effect is estimated. Thus, researchers face two competing considerations: they seek to efficiently estimate the direct effect of the treatment while minimizing bias due to spillover effects. Ex ante, we do not know whether spillovers occur; multilevel designs are essentially insurance against the risk of bias.

In order to illustrate this trade-off, we consider two research designs and explore the trade-offs inherent in each. In the first case, we randomly assign our experimental population into two equal-size groups, a treatment group (T) and a control group (C). The second case corresponds to the potential outcomes from a multilevel design. For ease of exposition, we assume that the outcomes are continuous, as opposed to binary, and char-

acterize each outcome in terms of its mean and variance. We calculate three test statistics and compare their statistical power. Without a multilevel design, we have randomly assigned the population into two groups, treatment and control. Suppose we have mean, variance, and sample size μ_T, σ_T^2, and n_T for the treatment group, and μ_C, σ_C^2, and n_C for the control group. Assume that $\sigma^2 = \sigma_T^2 = \sigma_C^2$ and $n_T = n_C = N/2$ where N represents the total sample. This assumes sampling from a superpopulation, or constant effects. This gives us a confidence interval of $CI = \mu_T - \mu_C \pm Z_\alpha \sqrt{\frac{4\sigma^2}{N}}$.

With a multilevel design, we have nine groups which have mean, variance, and sample size μ_{Gi}, σ_{Gi}^2, and n_{Gi}. We assume again that all groups have the same variance $\sigma_{Gi}^2 = \sigma_G^2$ and are of the same size, $n_{Gi} = \frac{N}{9}$. The test statistic for both the direct effects of treatment and spillover effects involve comparisons between two groups. For any given comparison of means $\mu_D = \mu_{Gi} - \mu_{Gj}$, we will therefore have a confidence interval of $CI = \mu_D \pm Z_\alpha \sqrt{\frac{18\sigma^2}{N}}$. In other words, the confidence intervals for our design may be substantially larger than they would be if we were to compare only two groups. This is a simple approximation, however, as multilevel designs induce dependencies between units that are not captured by this approximation. Modeling assumptions can yield efficiency improvements.

Empirical Application: Illinois 5th Congressional District Special Election 2009

Researchers who endeavor to study spillover effects experimentally must begin with a treatment that exerts an effect on those who receive it.[6] As noted in the previous section, the larger the direct effect, the greater the statistical power to detect spillovers. The same is true for sample size: an intervention that can be deployed on a grand scale is more likely to generate informative estimates of spillover effects. For these reasons, we used the social-pressure mailings developed by Gerber, Green, and Larimer (2008), which disclose whether a member of the household has voted in prior elections. Several large-scale follow-up experiments have demonstrated the robust turnout effects of this intervention (Gerber, Green, and Larimer 2010; Mann 2010) in low-salience elections.

---

[6] In principle, one could imagine a treatment that had no direct effect but still exerted a secondhand effect. No examples come to mind, however. Not surprisingly, prior studies that have used weak treatments in order to look for spillover effects (e.g., Ha and Karlan 2009) have failed to find them.

In addition to generating strong effects with minimal failure-to-treat problems (which would arise in the context of a phone or canvassing study), this type of mailing has a further advantage: its unusual message is plausibly the subject of conversation between housemates or neighbors.

It is also important to select an empirical context that seems conducive to the transmission of effects. As we indicated in our review of the large literature on political participation, interpersonal communication is widely thought to influence voter turnout in light of the predictive power of contextual factors such as friends' and neighbors' voting rates. In the dataset described below, spatial correlation in voting rates is quite powerful. Among households with two registered voters, turnout is just 12% when one's housemate abstains and 67% when one's housemate votes. Similarly, regressions that predict one's voter turnout based on the voter-turnout rate among others living in one's nine-digit zip code generate powerful and highly significant effects, even after controlling for a host of background factors including voter turnout in each previous election.[7] The purpose of our experiment is to assess whether the inferences suggested by nonexperimental evidence hold up when the voting rates of housemates and neighbors are manipulated exogenously.

In the final week prior to the April 2009 special election, approximately 60,000 individual residents of the Illinois 5th Congressional District[8] received a postcard in the mail reminding them to do their civic duty and vote in spring elections.[9] This postcard included a reminder about whether or not each individual had participated in the previous spring 2006 and spring 2008 elections and was patterned after the "self" mailings used by Green and Gerber (2008). An example of the postcard is shown in Figure 1. The postcard explicitly names a recipient and prints his or her voting history for the spring 2006 and spring 2008 elections, while leaving a line for the April 2009 election blank. We mailed postcards only to those residents who had been eligible to participate in both the 2006 and 2008 spring elections.

The April 7, 2009, election in the 5th Congressional District was unusual in that there was only a single elec-

toral contest on the ballot. Rahm Emanuel had previously held the seat but resigned to become the White House Chief of Staff. Twenty-four candidates competed in the March 3 special election primary, and three candidates emerged to compete in the April 7 special election: Mike Quigley (Democrat), Rosanna Pulido (Republican), and Matt Reichel (Green). Quigley defeated both candidates, winning 69.2% of the 44,138 votes. The contest received very little media coverage, and the Republican Party did very little campaigning for Pulido, who is a founder of the Illinois branch of the anti-illegal-immigration group known as Minutemen (Isenstadt 2009). The lack of competition is typical of elections for U.S. House of Representatives in districts where one party enjoys a lopsided advantage in terms of registered voters. The dearth of campaign communications is a plus from the standpoint of our experiment, as it means that our intervention and any interpersonal influence it generates are less likely to be obscured by competing communications.

Randomization and the Multilevel Design

Our experimental universe is a subset of all households in the 5th Congressional District. Individuals eligible to be part of the study were active voters with a permanent address who had registered before 2006. We consider eligible households to be households with between one and three registered voters. We then restricted the study to nine-digit zip codes containing at least two households with two eligible individuals and somewhere between 3 and 15 total households.[10] In our study we have 71,127 eligible individuals, 47,851 eligible households, and 4,897 eligible zip codes.

In order to allocate voters to experimental groups, the following procedure was used. We first conducted a randomization based upon each zip code. From the eligible zip codes, we selected a two-voter household at random, which we will call the core household. We then

---

[7] See Table A3.

[8] The 5th Congressional District is located north of the city of Chicago and borders Lake Michigan to the east. According to the U.S. Census in 2006, it has approximately 653,647 residents with a median income of $48,531. The district is 77.5% White, 23% Hispanic, 6.5% Asian, and 2.3% Black. The average Census block group within the district is 59% single-unit dwellings.

[9] We mailed the postcard April 2 and expected most residents to receive the postcard either April 3 or April 4.

[10] We rely upon an individual's nine-digit zip code as a proxy for their neighborhood. Nine-digit zip codes refer to the five-digit zip code plus a four-digit add-on number which identifies a compact geographic segment within the five-digit delivery area, such as a city block, office building, or individual high-volume receiver of mail (Maponics 2010). Our use of nine-digit zip codes is intended to capture very local neighborhoods, such as defined city blocks. In our study, each nine-digit zip code is comprised of roughly 15 eligible voters. Some nine-digit zip codes include the same apartment building, and the largest nine-digit zip code in our study is one city block, with a length of .2 miles. Begun in 1983, the purpose of the nine-digit zip codes is to aid efficient mail sorting and delivery (Grubesic 2006; Grubesic and Matisziw 2006). Across the participants within the study, the average distance between any two individuals within the same nine-digit zip code is 60 feet.

FIGURE 1 Sample Mailer

```
Dear Richard L Jensen:

DO YOUR CIVIC DUTY AND VOTE ON APRIL 7!

Why do so many people fail to vote?  We've been talking about
this problem for years, but it only seems to get worse --
especially when elections are held in the spring.

This year we're taking a different approach.  We're reminding
people that who votes is a matter of public record.  The chart
shows your name from the list of registered voters and whether
you voted in the last two spring elections.  The chart also
contains an empty space that we will fill in based on whether
you vote in the April 7th election.

DO YOUR CIVIC DUTY AND VOTE ON APRIL 7!

VOTER NAME                   Spring 2006  Spring 2008   April 7
RICHARD L JENSEN             Didn't Vote  Didn't Vote   _____



For more information: (517) 351-1975
Email: ETOV@Grebner.com

Practical Political Consulting
PO Box 6249
East Lansing MI 48826



                          RICHARD L JENSEN
                          3217 DABNEY ST
                          FRANKLIN PARK IL 60131-1503
```

chose 25% of core households to be assigned to control, and 75% were assigned to treatment. For zip codes where core households were assigned to control, all other households in their zip code were also assigned to control. For zip codes where core households were assigned to treatment, zip codes were assigned with equal probability into the following three conditions: First, with one-third probability, treatment core households were assigned to be the only household in their zip code receiving treatment. Second, with one-third probability, half of the other households in the zip code of the core treatment household were assigned to treatment. When there was an odd number of noncore households in the zip code, the number assigned to treatment was rounded down. (For example, with seven total households in the zip code, four would have been designated treatment households.) Finally, with one-third probability, all households in the zip code of the treatment core household were also assigned to treatment. Within households assigned to treatment, we randomly chose exactly one individual to receive treatment. Therefore, in one-person households, the full household was treated.

In two-person and three-person households, each person had a one-half and one-third probability of receiving the treatment, respectively.

After the election, we gathered official turnout records for each subject. Because some people moved or registered at a different address, 6,682 individuals were not found on the post-election voter rolls. Attrition, as expected, is not predicted by experimental assignment.[11] The analysis below focuses on the 64,445 individuals for whom we have valid outcome data (90% of our original sample).

In order to verify that the randomization procedure worked as planned, we conducted a series of multinomial

---

[11] We used randomization inference to test the sharp null hypothesis that treatment assignment has no effect on missingness. The test statistic was the F from a regression of attrition on indicator variables for each of the treatment conditions, controlling for the four-household strata. The p-value is 0.160. Assessing the interactions between treatment assignment and the covariates listed in Table 2 using randomization inference, we find no support for the hypothesis that the causes of attrition vary systematically across experimental groups. The p-value is 0.263.

TABLE 1 Treatment Assignment

	Full Sample		After Attrition	
	Control	Treatment	Control	Treatment
No Other HH Treated in Zip and Household Control	17,752	0	15,984	0
No Other HH Treated in Zip and Household Treatment	1,213	1,219	1,095	1,119
1 Other HH Treated in Zip and Household Control	15,308	0	13,874	0
Half of HH Treated in Zip and Household Control	8,032	0	7,279	0
Half of HH Treated in Zip and Household Treatment	3,451	6,296	3,129	5,179
All HH Treated in Zip and Household Treatment	5,793	12,063	5,313	10,933
Total	51,549	19,578	46,674	17,771

Note: Individuals are the unit of assignment that directly receives treatment.

TABLE 2 Outcome and Covariate Means

	Household Size			
	One Person	Two-Person Core	Two-Person Noncore	Three Person
Voted in Special 2009 Election	18	25	24	21
Voted in 2008 Primary	46	52	54	49
Voted in 2006 Primary	30	37	38	35
Voted in 2006 General	53	61	63	56
Voted in 2004 Primary	32	38	39	37
Voted in 2004 General	72	81	81	76
Voted in 2002 Primary	35	43	46	42
Voted in 2002 General	49	60	62	55
Voted in 2000 Primary	23	30	31	29
Voted in 2000 General	57	69	71	64

Note: Entries are the percentage of the subjects who voted in each election.

logistic regressions in which the nine assignment categories depicted in Table 1 were predicted using the list of covariates described in Table 2. This regression was performed on the postattrition sample within each of the four strata (one-person, two-person core, two-person noncore, and three-person households), resulting in four randomization checks. The joint significance of each chi-square test was assessed by calculating the sum of the chi-squares across the strata. Randomization inference was used to obtain p-values. In order to form the empirical sampling distribution of the test statistic under the null hypothesis, we simulated 10,000 randomizations. The results presented in Table 3 show the expected degree of balance. In each stratum and over all strata taken together, as expected, the likelihood-ratio test is nonsignificant (p > .05). Because the experimental groups are balanced with respect to past voting patterns, the regression estimates presented below are scarcely affected by whether votes cast in previous elections are introduced as covariates.

For ease of exposition, we use a linear probability model to estimate treatment and spillover effects. (The results are substantively unchanged when we use logistic regression.) As noted above, it is important to control for the strata within which individuals were randomized so that conditional on strata, each individual has an identical probability of receiving each treatment. For one-person households, for example, the model is

$$Y_i = \alpha + \beta \text{ Treatment}_i + \delta_1 \text{ 1 Other HH}_i$$
$$+ \delta_2 \text{ Half HH}_i + \delta_3 \text{ All HH}_i + X\Gamma + u_i, \quad (1)$$

where β represents the average effect of receiving the postcard oneself, and the δ_k denote the effects of residing in a zip code in which one, half, or all of the other households were treated.[12] The vector Γ represents fixed effects

[12] The variable 1 Other HH is an indicator variable scored 1 if an individual lives in a zip code in which just one other household is treated. Half HH and All HH variables refer to whether half or all of the other households in the zip code were treated.

TABLE 3 Balance by Household Size

1-person households	LR $\chi^2(36) = 42.48$	Prob $> \chi^2 = 0.585$	N = 25,002
2-person core households	LR $\chi^2(54) = 58.59$	Prob $> \chi^2 = 0.318$	N = 8,844
2-person noncore households	LR $\chi^2(54) = 64.62$	Prob $> \chi^2 = 0.433$	N = 21,250
3-person households	LR $\chi^2(54) = 48.05$	Prob $> \chi^2 = 0.928$	N = 9,349

Note: Each row reports the results of a multinomial logit regression testing whether the nine covariates included in Table 2 jointly predict assignment to treatment.
The dependent variable is the nine-category treatment assignment depicted in Table 1.
The results show no unexpected relationship between treatment assignment and past voting.
For two- and three-person households, there are seven assignment categories.
For one-person households, there are five assignment categories.

for each of the possible configurations of one-, two-, and three-person households within a zip code.[13]

In the case of three-person households, the regression model also includes a variable indicating whether another person within one's household is treated:

$$Y_i = \alpha' + \beta' \text{ Treatment}_i + \delta'_1 \text{ 1 Other HH}_i$$
$$+ \delta'_2 \text{ Half HH}_i + \delta'_3 \text{ All HH}_i$$
$$+ \gamma' \text{ Untreated in Treated HH} + X\Gamma' + u_i. \quad (2)$$

Here, the parameter γ' represents the average treatment effect of residing in a household where one of the subject's housemates received the treatment postcard.

Finally, among two-person households, the model is augmented with an indicator variable to distinguish between core households and noncore households, as these two types of two-person households were assigned to treatment with different probabilities.

$$Y_i = \alpha'' + \beta'' \text{ Treatment}_i + \delta''_1 \text{ 1 Other HH}_i$$
$$+ \delta''_2 \text{ Half HH}_i + \delta''_3 \text{ All HH}_i$$
$$+ \gamma'' \text{ Untreated in Treated HH}$$
$$+ \epsilon'' \text{ Core Household}_i + X\Gamma'' + u_i. \quad (3)$$

The cluster-randomized design complicates hypothesis testing and the estimation of confidence intervals. When a zip code, for example, is assigned to the control group, all of the individual voters are assigned to the control condition as a cluster. Clustered assignment also occurs when a multivoter household is assigned to the control group. When subjects are assigned to experimental conditions as clusters rather than individuals, conventional hypothesis tests tend to exaggerate the degree of statistical significance and conventional estimators tend

to underestimate the standard errors. Hypothesis tests are restored to proper size by estimating the sampling distribution using randomization inference. Under the sharp null hypothesis of no effect for any observation, all potential outcomes are observed empirically. The sampling distribution under the sharp null may be closely approximated by repeatedly reassigning the subjects to experimental groups and calculating the regression estimate.

Table 4 presents the statistical results within each of the experimental strata with confidence intervals reported based on randomization inference. The confidence intervals are estimated assuming constant effects equal to the estimated effect. In order to form the empirical sampling distribution of the test statistic, we simulated 10,000 randomizations using our randomization design. As noted above, each of the regression models includes individual treatment, whether another person in the household was treated,[14] and the extent of the zip-code-level treatment. The table also presents each regression with and without covariates, indicator variables indicating whether each individual voted in the previous nine elections. The inclusion of these covariates is optional; the estimator is consistent with or without them. The reason to include covariates is that by reducing disturbance variance, they may decrease the sampling variability of the estimated treatment effects. Both expectations are borne out in our data: including covariates has little effect on the estimates but slightly reduces their standard errors.[15]

Our attempt to measure spillover effects is predicated on the assumption that the postcard significantly increases the probability of turnout among those who receive it. Consistent with past experiments using a similar "self" mailer (Gerber, Green, and Larimer 2008, 2010;

---

[13] In this specification, a matrix of indicator variables (X) is introduced for each configuration of one-, two-, and three-voter households within each zip code. For example, all zip codes with exactly five one-voter households, three two-voter households, and three three-voter households would have the same fixed effect. There are roughly 300 unique configurations in our dataset. See Table 4.

[14] This variable is relevant only to individuals living in two- and three-voter households.

[15] Table 4 also includes fixed effects for each configuration of one-, two-, and three-person households within a zip code.

TABLE 4 Regression Estimates of Treatment and Spillover Effects

Household Size	One Person		Two Person		Three Person	
Individual	0.053	0.045	0.036	0.035	0.051	0.050
Treatment	(0.034, 0.071)	(0.028, 0.061)	(0.019, 0.052)	(0.02, 0.050)	(0.003, 0.097)	(0.006, 0.091)
1 Other HH	0.005	−0.001	0.005	0.006	−0.030	−0.021
Treated in Zip	(−0.012, 0.023)	(−0.018, 0.015)	(−0.016, 0.028)	(−0.015, 0.028)	(−0.067, 0.005)	(−0.054, 0.011)
Half of HH	−0.004	−0.008	0.006	0.008	−0.009	−0.011
Treated in Zip	(−0.023, 0.015)	(−0.025, 0.011)	(−0.013, 0.026)	(−0.011, 0.026)	(−0.051, 0.033)	(−0.049, 0.028)
All HH Treated	−0.010	−0.006	0.010	0.005	−0.033	−0.024
in Zip	(−0.035, 0.015)	(−0.029, 0.016)	(−0.012, 0.03)	(−0.014, 0.024)	(−0.090, 0.023)	(−0.076, 0.027)
Untreated in			0.012	0.012	0.020	0.021
Treated HH			(−0.005, 0.030)	(−0.003, 0.027)	(−0.027, 0.065)	(−0.021, 0.060)
Controls for Vote History	No	Yes	No	Yes	No	Yes
Strata Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	25,002	25,002	30,094	30,094	9,349	9,349
Clusters	4,782	4,782	13,683	13,683	2,294	2,294

Note: Design-based 95% confidence intervals in parentheses.

Mann 2010), the postcard increases turnout by approximately 5.3 percentage points among one-voter households, 3.6 percentage points among two-voter households, and 5.1 percentage points among three-voter households. Each of these estimates is significantly different from zero using confidence intervals based on randomization inference.[16] When we calculate a precision-weighted average of these estimates, we obtain a pooled estimate of 4.4 percentage points with a standard error of 0.6.[17] This pooled estimate is almost identical to the estimates reported in previous experiments, which also took place in low-salience elections. To put these effects in substantive perspective, note that the voting rates in the control group (those who received no mail and resided in zip codes where no mail was sent) vary from 16.4% to 22.2%. Thus, a 4.4 percentage-point effect represents between a 20% and 27% increase in turnout, which in the context of the voter-mobilization literature ranks as an unusually strong effect (Green and Gerber 2008).

Although the mailer had a clear effect on recipients, we find somewhat equivocal evidence of within-household spillovers. Nickerson (2008) finds that approx-

imately 60% of the effect of voter-mobilization messages is transmitted to those living in the same household. Considering only two- and three-person households, we estimate a precision-weighted average of the treatment effect to individuals of 3.8 with a standard error of 0.8. As applied to our data, Nickerson's finding implies that voters who reside in the same household as treated voters should vote at a 3.8 x 0.60 = 2.3 percentage point higher rate than the pure control group (which neither receives treatment nor resides in a household where others are treated). In fact, we find a pooled estimate of 1.3 percentage points with a 0.9 percentage-point standard error. Our best guess is that 34% of the direct effect of receiving a mailer is transmitted to housemates.[18] Sampling variability may explain why our estimates are somewhat lower than Nickerson's, but another possibility is that a conversation at the doorstep with a canvasser is more likely to prompt within-household conversation than our postcard.

We find little evidence of spillover effects across households within zip code. None of the estimated effects of high, medium, or low saturation of the zip code is significantly positive; instead, the estimates hover near zero. A precision-weighted average of the three

[16] In order to calculate confidence intervals, we created the full schedule of potential outcomes for each observation by assuming treatment effects equal to the estimates reported in Table 4.

[17] A test of heterogeneity fails to reject the null of equal effects across all three household sizes. $Q(2) = 1.9$, $p = 0.39$. When treatment effects are estimated using covariate adjustment, the precision-weighted average is 0.040 with a standard error of 0.005.

[18] Pooling the two- and three-person household-spillover estimates together and estimating the within-household spillover as a ratio of the direct effect, our estimates using randomization inference produce a 95% confidence interval of −5.56 to 59.75%. Because we are estimating a ratio, this interval is asymmetrically distributed around 50%.

high-saturation estimates is 0.0 percentage points with a standard error of 0.8 percentage points. The corresponding estimate for medium saturation is 0.0 with a standard error of 0.7 percentage points. The estimate for low saturation is also 0.0 with a standard error of 0.7. With a fair degree of precision, we find no indication that voters are affected indirectly by either the mobilization messages that their neighbors receive or the exogenous increase in their neighbors' propensity to vote that results from these mailers.

Prospects for Multilevel Experimentation

Experimentation in political science is motivated in large part by the desire to relax the strong assumptions that accompany observational research (Druckman et al. 2006). Nowhere is the contrast between experimental and observational approaches more stark than in the study of social influence. For decades, observational researchers have argued that the influence of social networks is demonstrated by the correlation between friends', neighbors', or coworkers' attitudes and behaviors. Concerned that unobserved factors other than social influence might account for these correlations, experimental researchers endeavor to introduce exogenous messages in an effort to study the extent to which they ramify through existing networks.

Multilevel experiments are a logical extension of this push to relax assumptions. Rather than assume that experimental subjects respond only to the messages directed at them, researchers employing multilevel designs allow for the possibility that messages are transmitted indirectly from person to person. Given the vast number of social pathways through which messages could be transmitted, multilevel designs inevitably rely on a simplified representation of social transmission. In our empirical application, for example, we considered only two pathways: within household and within zip code. Yet even this relatively circumscribed multilevel design generates useful insights. Although the direct effect of the treatment is strong,[19] and roughly half of this influence is transmitted to housemates, we find no support for the hypothesis that effects are transmitted between neighboring households. Despite the strong correlation among neighbors'

probabilities of voting, voter-mobilization messages do not appear to exert secondhand effects on neighbors.

Although it is tempting to judge the experimental literature on social influence according to whether researchers find evidence of "interesting" spillovers, the importance of null findings should not be discounted. Given the vast number of conceivable pathways through which treatment effects could be transmitted, it is extremely helpful to have empirical grounds for ruling out certain kinds of SUTVA violations. The point of using a multilevel design is to estimate additional parameters; if those parameters are estimated with precision and discovered to be close to zero, multilevel design has done its job, and researchers can direct their attention to other, more pressing threats to internal validity.

The next frontier for multilevel experiments designed to assess interpersonal communication is to apply experimental designs to well-documented social networks. Unlike the current study, which uses place of residence as a proxy for network proximity and frequency of interaction, future experiments may build on detailed information about the nature and frequency of interpersonal contact, whether gathered via surveys, ethnographies, or social networking websites. The community defined by the social networking database comprises the experimental sample, and the researcher directs experimental treatments to randomly selected individuals and groups. By allowing the researcher to assess whether treatments diffuse at different rates according to centrality of the receiver within the social network, this type of design allows for a more nuanced investigation of social influence and SUTVA more generally.

Multilevel designs are potentially useful in a wide array of substantive domains outside the field of political behavior. One especially fruitful set of applications involves monitoring and deterrence. For example, in an effort to deter electoral fraud, international election observers visit randomly selected polling stations (Hyde 2010). These visits are intended to generate spillover effects; the monitors seek to signal their surveillance of the polling stations they visit as well as polling stations in the surrounding area. It would be relatively easy to craft a multilevel design in which monitors vary the proportion of polling places that are monitored in each town and vary the proportion of towns that are monitored in each region. The same general approach could be applied in a variety of other experimental monitoring efforts that seek to deter embezzlement by local government officials (Silva 2010) or to promote tax compliance by citizens (Felner, Sausgruber, and Traxler 2009). Another set of promising applications concerns what might be called "policy diffusion." Spillovers from research and

---

[19] Additional analysis in Table A2 indicates that the direct treatment effect interacts strongly with voting in prior elections; the treatment effect is especially large among those who regularly vote in primary elections. The interactive model, whether estimated using regression or probit, shows equivocal evidence of within-household spillovers and no neighborhood spillovers.

development (Audretsch and Feldman 1996), the institution of new environmental policies (Sheely 2010), and the elevation of women to local leadership positions (Beaman et al. 2009) are instances in which interventions occurring at one location may have repercussions for policy outcomes in nearby areas. The implementation of multilevel designs may be feasible in situations where experimental interventions are to be rolled out in a variety of locations, as in Sheely's study of environmental policies that were randomly introduced in 36 Kenyan villages. By varying the spatial arrangement of experimental interventions, multilevel design fosters the systematic investigation of whether an intervention in one location influences policy adoption nearby. In sum, the range of phenomena that are potentially affected by spillovers, displacement, signaling, and social comparison is quite broad. Multilevel experiments represent a useful design-based strategy for investigating these phenomena.

Appendix

TABLE A1   Voting Rates by Zip Code, Household, and Individual Treatment Assignment

Assignment	3-Person HH	2-Person Noncore	2-Person Core	1-Person HH	N
$\hat{y}_{000}$	21.41 (477/2,228)	21.55 (1,150/5,337)	25.20 (555/2,202)	16.42 (1,021/6,217)	15,984
$\hat{y}_{low00}$	18.84 (452/2,399)	22.88 (1,220/5,332)	—	17.14 (1,053/6,143)	13,874
$\hat{y}_{001}$	—	—	27.79 (311/1,119)	—	1,119
$\hat{y}_{010}$	—	—	22.19 (243/1,095)	—	1,095
$\hat{y}_{.500}$	21.21 (255/1,202)	22.85 (631/2,761)	—	15.86 (526/3,316)	7,279
$\hat{y}_{.501}$	25.19 (99/393)	25.59 (324/1,266)	26.91 (299/1,111)	21.02 (620/2,949)	5,719
$\hat{y}_{.510}$	22.69 (179/789)	24.40 (307/1,258)	25.23 (273/1,082)	—	3,129
$\hat{y}_{101}$	25.00 (196/784)	27.04 (714/2,641)	25.99 (294/1,131)	20.64 (1,316/6,377)	10,933
$\hat{y}_{110}$	21.49 (334/1,554)	25.08 (666/2,655)	24.09 (266/1,104)	—	5,313

Note: We send mail to at most one randomly selected member of each household, so we never observe V_{z11} outcomes.
The first digit in each voting rate refers to zip code; the second, to household; and the third, to individual.
The "low" designation indicates that just one other household in the zip code receives treatment.

TABLE A2   OLS Estimates of Direct and Spillover Treatment Effects, by Household Size and Turnout in Prior Elections

Variable	Voted in Both Previous	Voted in One Previous	Voted in None
One-Person Household			
Individual Treatment	.1228* (.0242)	.0168 (.0170)	.0176* (.0084)
1 Other HH Treated in Zip	.0303 (.0195)	−0.010 (.0141)	−.0043 (.0071)
Half of HH Treated in Zip	−.0249 (.0221)	.0090 (.0161)	−.0052 (.0078)

(Continued)

TABLE A2 (Continued)

Variable	Voted in Both Previous	Voted in One Previous	Voted in None
All HH Treated in Zip	−.0343	.0259	−.0056
	(.0310)	(.0227)	(.0111)
Constant	.3910**	.1521**	0.0549**
	(.0140)	(.0010)	(.0050)
N	6,264	7,120	11,618
R-squared	.00	.00	.00
Clusters	3,113	3,469	4,170
Two-Person Household			
Individual Treatment	.0614**	.0312*	.0074
	(.0211)	(.0178)	(.0101)
1 Other HH Treated in Zip	.0306	.0165	−.0085
	(.0196)	(.0161)	(.0099)
Half of HH Treated in Zip	.0052	.0235	−.0006
	(.0192)	(.0160)	(.0103)
All HH Treated in Zip	.0156	.0135	.0033
	(.0238)	(.0194)	(.0123)
Untreated in Treated HH	.0305	−.0147	.0095
	(.0210)	(.0174)	(.0109)
Core Household	.0266*	.0284*	.0017
	(.0160)	(.0137)	(.0081)
Constant	.4266**	.1774**	.0809**
	(.0128)	(.0102)	(.0068)
N	9,596	8,884	11,614
R-squared	.00	.00	.00
Clusters	5,784	5,659	6,939
Three-Person Household			
Individual Treatment	.0573	.1213**	−.0028
	(.0490)	(.0376)	(.0207)
1 Other HH Treated in Zip	−.0657*	−.0154	.0054
	(.0336)	(.0270)	(.0141)
Half of HH Treated in Zip	−.0033	−.0362	.0158
	(.0407)	(.0295)	(.0166)
All HH Treated in Zip	−.0284	−.0295	.0321
	(.0566)	(.0441)	(.0239)
Untreated in Treated HH	.0269	.0546	−.0254
	(.0463)	(.0347)	(.0190)
Constant	.4634**	.1761**	0.0657**
	(.0244)	(.0199)	(.0098)
N	2,725	2,561	4,063
R-squared	.01	.01	.00
Clusters	1,333	1,405	1,803

Note: Previous voting status refers to the two elections listed on the postcard.
Robust cluster standard errors in parentheses, where the cluster is the household.
One-tailed tests; *significant at 5%, **significant at 1%.

TABLE A3 **Estimates of Spillover Effects Based on a Nonexperimental Identification Strategy That Treats Household and Neighborhood Voting Rates as Exogenous, for Core Two-Person Households**

Variable	Model 1	Model 2
Individual Treatment	.036**	.035**
	(.008)	(.007)
9-Digit Zip Code Voting Rate	.310**	.381**
	(.022)	(.022)
Voter Turnout of Other	.495**	.427**
Household Member	(.010)	(.010)
Male		.034**
		(.007)
White		.020**
		(.008)
Democratic Registration		.016
		(.009)
Number of Votes Cast		.045**
Since 2000		(.002)
Age		.002**
		(.000)
Constant	.061**	−.179**
	(.006)	(.015)
N	8844	8753
Adjusted R-squared	.304	.382

Note: One-tailed tests; *significant at 5%, **significant at 1%. Zip code voting rates exclude the subject's own turnout.
Difference in number of observations in second model reflects missing data in the additional covariates.

References

Audretsch, David B., and Maryann P. Feldman. 1996. "R & D Spillovers and the Geography of Innovation and Production." *American Economic Review* 86(3): 630–40.

Bednar, J., Y. Chen, T. X. Liu, and S. Page. 2010. "Behavioral Spillovers and Cognitive Load in Multiple Games: An Experimental Study." Working paper, University of Michigan.

Beaman, Lori, Raghabendra Chattopadhyay, Esther Duflo, Rohini Pande, and Petia Topalova. 2009. "Powerful Women: Does Exposure Reduce Bias?" *Quarterly Journal of Economics* 124(4): 1497–1540.

Bolton, Charles. 1972. "Alienation and Action: A Study of Peace Group Members." *American Journal of Sociology* 78 (November): 537–61.

Briet, Martien, Bert Klandermans, and Frederike Kroon. 1987. "How Women Became Involved in the Women's Movement of the Netherlands." In *The Women's Movements of the United States and Western Europe: Consciousness, Political Opportunities, and Public Policy*, ed. Mary Katzen-

stein and Carol Mueller. Philadelphia, PA: Temple University Press, 44–63.

Cho, Wendy K. Tam, James G. Gimpel, and Joshua J. Dyck. 2006. "Residential Concentration, Political Socialization, and Voter Turnout." *Journal of Politics* 68(1): 156–67.

Cohen, Cathy J., and Michael C. Dawson. 1993. "Neighborhood Poverty and African American Politics." *American Political Science Review* 87(2): 286–302.

David, Olivier, and Rob A. Kempton. 1996. "Designs for Interference." *Biometrics* 52(2): 597–606.

Druckman, Jamie D., Donald P. Green, James H. Kuklinski, and Arthur Lupia. 2006. "The Growth and Development of Experimental Research Political Science." *American Political Science Review* 100(4): 627–35.

Duflo, Esther, and Emmanuel Saez. 2003. "The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment." *Quarterly Journal of Economics* 118(3): 815–42.

Fellner, Gerlinde, Rupert Sausgruber, and Christian Traxler. 2009. "Testing Enforcement Strategies in the Field: Legal Threat, Moral Appeal and Social Information." CESifo Working Paper Series No. 2787.

Gerber, Alan S., Donald P. Green, and Christopher W. Larimer. 2008. "Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment." *American Political Science Review* 102(February): 33–48.

Gerber, Alan S., Donald P. Green, and Christopher W. Larimer. 2010. "An Experiment Testing the Relative Effectiveness of Encouraging Voter Participation by Inducing Feelings of Pride or Shame." *Political Behavior* 32(3): 409–22.

Gerlach, Luther, and Virginia Hine. 1970. *People, Power and Change: Movements of Social Transformation*. Indianapolis, IN: Bobbs-Merrill.

Gimpel, James G., Joshua J. Dyck, and Daron R. Shaw. 2004. "Registrants, Voters, and Turnout Variability across Neighborhoods." *Political Behavior* 26(4): 343–75.

Green, Donald P., and Alan S. Gerber. 2008. *Get Out the Vote: How to Increase Voter Turnout*. 2nd ed. Washington, DC: Brookings Institution Press.

Grosser, Jens, and Arthur Schram. 2006. "Neighborhood Information Exchange and Voter Participation: An Experimental Study." *American Political Science Review* 100(2): 235–48.

Grubesic, Tony H. 2008. "Zip Codes and Spatial Analysis: Problems and Prospects." *Socio-Economic Planning Sciences* 42(2): 129–49.

Grubesic, Tony H., and Timothy C. Matisziw. 2006. "On the Use of ZIP Codes and ZIP Code Tabulation Areas (ZCTAs) for the Spatial Analysis of Epidemiological Data." *International Journal of Health Geographics* 5(58).

Ha, Shang E., and Dean S. Karlan. 2009. "Get-Out-the-Vote Phone Calls: Does Quality Matter?" *American Politics Research* 37(2): 353–69.

Haber, Michael. 1999. "Estimation of the Direct and Indirect Effects of Vaccination." *Statistics in Medicine* 18: 2101–9.

Halloran, M. Elizabeth, and Claudio J. Struchiner. 1991. "Study Designs for Dependent Happenings." *Epidemiology* 2(5): 331–38.

Halloran, M. Elizabeth, and Claudio J. Struchiner. 1995. "Causal Inference in Infectious Diseases." *Epidemiology* 6(2): 142–51.

Hong, Guanglei, and Stephen W. Raudenbush. 2006. "Evaluating Kindergarten Retention Policy: A Case Study of Causal Inference for Multilevel Observational Data." *Journal of the American Statistical Association* 101(475): 901–11.

Huckfeldt, R. Robert. 1979. "Political Participation and the Neighborhood Social Context." *American Journal of Political Science* 23(3): 579–92.

Hudgens, Michael G., and M. Elizabeth Halloran. 2006. "Causal Vaccine Effects on Binary Postinfection Outcomes." *Journal of the American Statistical Association* 101(473): 51–65.

Hudgens, Michael G., and M. Elizabeth Halloran. 2008. "Toward Causal Inference with Interference." *Journal of the American Statistical Association* 103(482): 832–43.

Hyde, Susan. 2011. "Catch Us If You Can: International Election Monitoring and Norm Diffusion." *American Journal of Political Science* 55(2): 356–69.

Imbens, Guido, Joshua D. Angrist, and Donald Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of Econometrics* 71(12): 145–60.

Isenstadt, Alex. 2009. "Republicans' Outlook Remains Bleak." *Chicago Tribune*, May 2.

Knoke, David. 1990. "Networks of Political Action: Toward Theory Construction." *Social Forces* 68(4): 1041–63.

Mann, Christopher B. 2010. "Is There Backlash to Social Pressure? A Large-Scale Field Experiment on Voter Mobilization." *Political Behavior* 32 (September): 387–407.

Manski, Charles. 1993. "Identification of Exogenous Social Effects: The Reflection Problem." *Review of Economic Studies* 60: 531–42.

Maponics. 2010. "ZIP Code Maps—Frequently Asked Questions." Accessed April 5, 2012. http://www.maponics.com/ZIP_Code_Maps/ZIP_Code_FAQ/zip_code_faq.html.

McAdam, Doug. 1986. "Recruitment to High-Risk Activism: The Case of Freedom Summer." *American Journal of Sociology* 92(1): 64–90.

McAdam, Doug, and Ronnelle Paulsen. 1993. "Specifying the Relationship between Social Ties and Activism." *American Journal of Sociology* 99(3): 640–67.

McClurg, Scott D. 2003. "Social Networks and Political Participation: The Role of Social Interaction in Explaining Political Participation." *Political Research Quarterly* 56(4): 449–64.

Miguel, Edward, and Michael Kremer. 2004. "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities." *Econometrica* 72(1): 159–217.

Nickerson, David. 2008. "Is Voting Contagious? Evidence from Two Field Experiments." *American Political Science Review* 102(February): 49–57.

Ratcliffe, J., T. Taniguchi, E. R. Groff, and J. Wood. 2011. "The Philadelphia Foot Patrol Experiment: A Randomized Controlled Trial of Police Effectiveness in Violent Crime Hotspots." *Criminology* 49(3): 795–831.

Rosenbaum, Paul R. 2007. "Interference between Units in Randomized Experiments." *Journal of the American Statistical Association* 102(477): 191–201.

Rubin, Donald B. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 55(5): 688–701.

Rubin, Donald B. 1978. "Bayesian Inference for Causal Effects: The Role of Randomization." *The Annals of Statistics* 6(1): 34–58.

Rubin, Donald B. 1980. "Randomization Analysis of Experimental Data: The Fisher Randomization Test Comment." *Journal of the American Statistical Association* 57(371): 591–93.

Rubin, Donald. 1986. "Which Ifs Have Causal Answers? Discussion of Holland's 'Statistics and Causal Inference'." *Journal of the American Statistical Association* 81(396): 961–62.

Sherman, L., and D. Weisburd. 1995. "General Deterrent Effects of Police Patrol in Crime Hot Spots: A Randomized Controlled Trial." *Justice Quarterly* 12: 625–48.

Sheely, Ryan. 2009. "The Role of Institutions in Providing Public Goods and Preventing Public Bads: Evidence from a Public Sanitation Field Experiment in Rural Kenya." Working paper.

Silva, Pedro. 2010. "Learning to Fear the Inspector-General: Measuring Spillovers from Anti-Corrupt Policies." Working paper.

Sobel, Michael E. 2006. "What Do Randomized Studies of Housing Mobility Demonstrate? Causal Inference in the Face of Interference." *Journal of the American Statistical Association* 101(476): 1398–1408.