

The Role of Call Quality in Voter Mobilization: Implications for Electoral Outcomes and Experimental Design

Christopher B. Mann · Casey A. Klofstad

© Springer Science+Business Media New York 2013

Abstract We demonstrate the centrality of high quality personal interactions for successfully overcoming the collective action problem of voter mobilization, and highlight the need for attention to treatment quality before making substantive inferences from field experiments. We exploit natural variation in the quality of voter mobilization phone calls across call centers to examine how call quality affects voter mobilization in a large-scale field experiment conducted during the 2010 Election. High quality calls (from call centers specializing in calling related to politics) produced significant increases in turnout. In contrast, low quality calls (from multi-purpose commercial call centers) failed to increase turnout. Furthermore, we offer caution about using higher contact rates as an indication of delivery quality. Our treatment conditions with higher contact rates had no impact on turnout, suggesting an unfavorable trade-off between quantity of contacts and call quality.

Keywords Field experiment · Voter mobilization · Causal inference · Experimental design · Mobilization calls · House effects

Electronic supplementary material The online version of this article (doi:[10.1007/s11109-013-9264-y](https://doi.org/10.1007/s11109-013-9264-y)) contains supplementary material, which is available to authorized users.

C. B. Mann
Department of Political Science, Louisiana State University, Baton Rouge, LA 70803, USA

C. B. Mann (✉)
Manship School of Mass Communication, Journalism Building, Baton Rouge, LA 70803, USA
e-mail: cmann13@lsu.edu

C. A. Klofstad
Department of Political Science, University of Miami, 1300 Campo Sano Avenue, Coral Gables, FL 33416, USA
e-mail: c.klofstad@miami.edu

In a democratic society, participation in elections ensures representation of the citizenry and enhances the legitimacy of the government (Dahl 1956). However, encouraging this collective action is difficult (Olson 1965). Consequently, scholars have paid considerable attention to how citizens can be mobilized, particularly in the United States where turnout rates lag behind other advanced democracies. Campaigns frequently use phone calls to try to mobilize voters, but why would we expect a phone conversation with a complete stranger overcome the collective action problem in voting?

Examining the current array of “get out the vote” (GOTV) field experiments reveals that some phone calls significantly increase turnout, while others have no effect (Green and Gerber 2008, pp. 74–96). Numerous field experiments over the last decade have examined whether the messages used in voter mobilization phone calls explain this variation. However, scholars have found little evidence to suggest that message influences the effectiveness of these calls (Green and Gerber 2008, pp. 74–96). By comparison, the quality of mobilization calls has been under-studied. While the literature has established an important distinction between hasty impersonal calls and personal chatty calls (Nickerson 2007), there has been little exploration of the potential for a wider range of variation in call quality beyond this personal/impersonal categorization. Given the extensive use of phone calls in field experiments on political behavior and political campaigns, a more nuanced understanding of call delivery quality is needed.

In this study, we demonstrate that variation in other aspects of call quality significantly influence the effectiveness of voter mobilization phone calls. We used a large-scale field experiment conducted in eleven states during the 2010 General Election ($N = 637,484$). The experiment randomly assigned registered voters to four levels of call quality while holding all else constant about the mobilization contact.

Our findings make five contributions. First, we demonstrate that call quality has a significant influence on the effectiveness of voter mobilization calls. Second, our findings about call quality provide additional insights into what makes voter mobilization communication from strangers effective at getting people out to vote. Third, we highlight the importance of attending to the quality of treatment delivery before reaching conclusions about theoretical propositions tested in field experiments. More specifically, scholars must account for the quality of treatment delivery before drawing inferences about message or other mechanisms from field experiments. Fourth, we caution scholars about potential dangers of emphasizing contact rates to attempt to increase statistical power. In our experiment, higher contact rates were correlated with weaker treatments, and thus lower statistical power to detect true treatment effects. Finally, we provide practical guidance on the type of call centers most likely to increase participation in elections.

The paper proceeds as follows: In the next section, we discuss voter mobilization call quality within the context of the extant literature. We consider the utility and importance of our definition of call quality relative to focusing on contact rate, a measure frequently used by scholars and political professionals. Then we discuss why attending to the quality of treatment delivery is essential to our understanding

of voting and other political behavior when using field experiments. We state the hypotheses tested in the field experiment, before outlining the research design of the field experiment. We present the results of the field experiment and conclude with a discussion of our findings.

Assessing Call Quality

Phone calls are a staple of voter mobilization. Scholars use field experiments to examine whether these calls increase participation. These studies deliver stimuli to registered voters in the form of phone calls, and measure the effects on participation by examining individual level voter turnout records maintained by local election agencies. Early field experiments found little evidence that phone calls increased turnout (Gerber and Green 2000, 2001, 2005a, b). However, Gerber and Green (2001, p. 77) suggested call quality may influence the success of mobilization phone calls, noting “[t]he calls sounded as though they were made by a professional firm rather than local volunteers or neighbors. The telephone scripts were generally delivered competently, but sometimes hastily or mechanically”.

The most significant insight on call quality to date came when Nickerson found large effects from voter mobilization calls made by volunteers (Nickerson 2006, 2007). Volunteers naturally take a conversational approach to calls as they interact with a respondent. In Nickerson’s (2007) comparison of mobilization calls by paid callers and volunteers, calls with a conversational tone significantly increase voter participation. Paid callers can be nearly as effective when coached to be conversational, and the effectiveness of volunteers disappears when they are pressured into hasty and mechanical delivery. A field experiment by Ha and Karlan (2009) confirms the effect of using a conversational style call by varying “interactiveness” of the script (the number of questions asked of respondents) using calls made by a single commercial phone bank.

These findings suggest that scholars and political professionals should use volunteer phone banks. However, most mobilization calls in the United States, and research about those calls, rely on professional call centers (albeit with instructions for the callers to be conversational). Additionally, replications of previously successful field experiments using professional call centers have found wide variation in effects on turnout, including null effects (Green and Gerber 2008; Ha and Karlan 2009; Garcia Bedolla et al. 2012). Although electoral context and other factors play a role in these differences, variation in call quality is a major contributor as well. Our experiment directly tests the degree to which variation in call quality across different types of professional call centers contributes to variation in turnout effects.

Our approach is also motivated by the survey research literature on the influence of call center procedures on survey responses (“house effects”). Survey research is quite different from mobilizing voters. Nonetheless, research showing that call quality has notable effects on survey results suggests that we should ask similar questions about the impacts of call quality on voter mobilization treatments.

The survey research literature on house effects shows different survey firms can produce different survey results, even when they simultaneously conduct polls with similar questions and questionnaire design at the same time (for a summary, see Leahey 2008 and Weisberg 2005, pp. 299–303). This variation is linked to different operating procedures across firms (Smith 1978, 1982; Porter 1995; Viterna and Maynard 2002). For example, Porter (1995) showed that commercial polling firms tend to have stricter protocols because they employ less experienced interviewers, while interviewers at academic polling centers are allowed to act more autonomously because they are well-trained. These different approaches lead to variation in the survey responses that are collected. For example, rigid operating procedures may not be nimble enough to handle situations, such as responding to an answer of “don’t know” that is more easily addressed by seasoned interviewers who are allowed to act extemporaneously. Viterna and Maynard (2002) found variation in the strictness of operating procedures between academic polling centers, on a variety of dimensions such as the pace and tone of interviews, all of which can affect the responses provided to interviewers.¹

The survey research literature also has a clear standard of call quality. A quality call is one that produces unbiased responses to survey questions. For voter mobilization research we define call quality as credible and sincere delivery of the mobilization message. In both cases, the standard is concerned with how subjects respond to the calls. Poor quality survey calls will lead to biased reports of attitudes by respondents. For example, poor rapport between interviewer and respondent might prompt the respondent to reply in haste to complete the conversation more quickly, or even hang up the phone before the interview is complete. Poor quality in mobilization calls will bias voter response to mobilization messages. For example, poor rapport between a caller and potential voter will not cause the subject to change his or her voting behavior.

In the experiment, we leverage natural variation in the operating practices of call centers that make voter mobilization phone calls. Different business models lead call centers to employ different types of callers and manage calling in different ways. The varying operating practices are expected to produce variation in call quality. We expected higher call quality from call centers directly operated by firms specializing in politically-related calls, and from call centers directly operated by firms specializing in telephone fundraising for political and non-profit organizations. One example of each is included in the experiment (treatments A and B). Since their business is specialized, these in-house call centers have callers who are more likely to be interested in politics. This interest and experience making calls about politics is expected to lead to more credible and sincere delivery of mobilization messages.

¹ The house effects literature largely focuses on how the training of interviewers influences how they interact with respondents, and what influence that has on the data collected (Leahey 2008; Smith 1978, 1982; Viterna and Maynard 2002). For example, highly trained and well-monitored interviewers are typically given more latitude to engage the respondent in a more conversational tone, and to act extemporaneously though the use probes and feedback to encourage the respondent. This approach can increase the quality of the data collected by reducing the amount of missing data due to breakoffs, refusals, and “do not know” responses. Commercial polling firms tend to give less latitude to interviewers, and have more standardized procedures, than academic polling centers.

Conversely, calls that are out-sourced to multi-purpose commercial call centers are expected to have lower quality. Hiring practices are unlikely to consider political interest of potential callers, and standard calling procedures are designed for the commercial calls that are the vast majority of their business. Therefore, these out-sourced calls are less likely to deliver credible and sincere calls about voter mobilization. Since this is the most common business model in the industry, two exemplars are included in the experiment to account for potential variation between out-sourcing firms (treatments C1 and C2).

Subjects were randomly assigned to be called by one of these four firms, or to a control group. The script was identical across all four treatments to isolate the effects of call quality. The outcome of interest in the experiment is a comparison of the efficacy of each treatment at increasing voter turnout.

Distinction Between Call Quality and Contact Rate

In contrast to our standard for call quality, political professionals often use contact rate as an indicator of quality for voter mobilization calls. They assume that reaching more people indicates high quality execution of the mobilization calling program. This assumption is similar to the frequent use of contact rate to assess the quality of data collected by surveys. However, a growing literature suggests that increasing contact rate does not necessarily decrease error in surveys (e.g., Keeter et al. 2006; Olson and Bilgen 2011), and may actually increase survey error (Atkeson et al. 2011).

The assumption that higher contact rates indicate higher quality voter mobilization calls is a testable proposition. In the analysis we compare the contact rates from the different call centers to impact on voter turnout to see if higher contact rates are correlated with larger effects on voter mobilization. We take the expectation of a positive correlation between contact rate and mobilization effect as the conventional hypothesis about call quality. A negative relationship between contact rates and mobilization effect rejects this hypothesis. The most likely mechanism for a negative relationship is that higher contact rates are achieved at the expense of call delivery quality. The potential benefit of reaching more people is offset by less effect from each call when contact rates are valued over delivery quality.

If subject responsiveness to mobilization is correlated with how easily they are contacted, one might argue that variation in treatment responsiveness is an alternative explanation for the relationship between contact rate and voter turnout. That is, the efficacy of the mobilization program changes as contact rate increases because individuals who are more difficult to contact have a different propensity to be mobilized by the treatment. We contend, however, that heterogeneity in treatment responsiveness is not an alternative explanation for the relationship between contact rate and turnout for four reasons.

First, consider the relationship between mobilization and contact rate if treatment responsiveness is constant or the most treatment-responsive subjects are also the most difficult to contact. Under these conditions the positive marginal impact on

mobilization will increase as contact rate increases—and in the latter case this marginal increase will accelerate as contact rates increase (i.e., looking something like an exponential effect). In these cases where mobilization increases as contact rates increase, we would still accept the conventional hypothesis that contact rate is indicative of call quality.

Second, if treatment-responsive individuals are the easiest to reach, the marginal impact on mobilization will decrease as contact rate increase (i.e., looking something like a logarithmic effect). However, this possibility of diminishing returns from a higher contact rate also exhibits a positive relationship between contact rate and mobilization rate as long as treatment responsiveness is greater than zero. In this case we would also accept the conventional hypothesis that contact rate is indicative of call quality.

Third, treatment-responsiveness would create a negative relationship between contact rate and voter mobilization only if subjects who are more difficult to reach respond negatively to the mobilization effort. That is, *de*-mobilization by a treatment intended to increase participation. While plenty of mobilization experiments report null effects, there is no evidence from the extensive array of voter mobilization field experiments showing significant demobilization effects, much less negative treatment-responsiveness among more difficult to reach individuals when a treatment has caused a positive effect among easier to reach voters.

Fourth, a null relationship between contact rate and voter mobilization is also grounds for rejecting the conventional hypothesis. However, a null effect may occur for at least two other reasons: the call delivery quality was entirely inadequate for all contacts, or the treatment message was ineffectual. We return to these dual possible explanations in the next section.

Role of Treatment Quality When Testing Voter Mobilization Hypotheses

The question of call quality raises an important methodological concern about interpreting observed treatment effects in field experiments. In principle, researchers design experiments to perfectly measure the effect of a treatment. In the experiment, T is a dichotomous condition randomly assigned within the experimental population (the subject is either treated or untreated). Hypothesis testing uses the parameter Δ , the true effect of the treatment (treatment effect = $\Delta \times T$). In practice, however, the treatment is not delivered perfectly. Thus, the actual treatment must be discounted by imperfections of implementation (ζ). Thus the observed effect (δ) of a treatment has two components, the true effect of the treatment (Δ) and the quality of treatment delivery (ζ), such that,

$$\delta = \Delta \times \zeta, \quad (1)$$

where $0 \leq \zeta \leq 1$, with 0 indicating such poor quality that no element of the ideal treatment is present, and 1 indicating the ideal treatment. As the quality of treatment delivery increases, δ becomes a more accurate measurement of the true Δ . Conversely, as the quality of treatment delivery declines, testing hypotheses about Δ is confounded by the inability to accurately distinguish Δ from ζ .

In voter mobilization field experiments, the treatments take the form of a phone call, mailing, face-to-face canvassing visit, or other communication. While scholars usually take great pains over the wording of treatments to capture the concept under investigation, the quality of treatment delivery (ζ) is often treated as a nuisance, an external validity issue, or simply ignored. However, before making substantive inferences about treatment effects in voter mobilization experiments—particularly when the experiment fails to reject the null hypothesis of no treatment effect—researchers must attend to quality of treatment delivery.

The literature on phone calls for voter mobilization illustrates why treatment delivery should be a concern. As noted above, early experiments on voter mobilization phone calls found no evidence of increased turnout. Nonetheless, political scientists continued to research variations in phone call treatments largely because political operatives were convinced of their effects. Without this pressure from political professionals, political scientists might have erroneously concluded that phone calls had little or no effect. Instead, scholars now consider phone calls second only to face-to-face canvassing as the most effective mode of mobilizing registered voters (Green and Gerber 2008).

We relied on natural variation between call center operating models for real-world differences in treatment delivery quality (ζ). We held constant the phone script across our experimental conditions, and therefore the true effect of this particular treatment (Δ) was also constant. We based the phone script on previous experiments that increased turnout, so we assume $\Delta > 0$. The results below replicated the increase in turnout ($\Delta > 0$) with high-quality phone calls. In contrast, the treatment failed to generate any effect with lower quality calls made by multi-purpose call centers, the types of phone calls most commonly used for voter mobilization and research on voter mobilization. Any experiment testing hypotheses about the wording, timing, or other variations in phone call treatments could not draw meaningful inferences when using these ineffectual low quality calls. Therefore, our findings highlight why researchers must pay attention to quality of treatment delivery before reaching conclusions about theoretical propositions.

Hypotheses

We test the following hypotheses about the quality of treatment delivery:

Hypothesis 1 High-quality calls will be more effective at causing an increase in voter turnout. More specifically, calls delivered from call centers more likely to have callers interested, skilled, and experienced in political calls will be more effective at increasing turnout.

Hypothesis 2 High contact rates are an indication of call delivery quality for voter mobilization calls. That is, higher contact rates will be associated with larger increases in voter turnout.

Research Design

We utilized a large-scale field experiment using phone calls to encourage turnout among registered voters in eleven states during the 2010 General Election ($N = 637,484$).² The calls began on the Friday prior to Election Day (October 29, 2010) and continued until Election Day (Tuesday, November 2, 2010). The experiment was conducted in partnership with a non-partisan organization seeking to increase voter turnout.³ Data for the experiment was provided by Catalist LLC, a firm specializing in individual-level voter data.

Random Assignment

Registered voters were randomly assigned to five conditions: a control group (no attempted phone calls), or voter mobilization calls by one of four call centers selected by our partner organization. We limited the experimental population to households with only one registered voter selected by our partner organization to avoid violating statistical assumptions about causal inference due to the correlation in behavior of co-habitants (Sinclair et al. 2012; Klostad et al. 2011; Nickerson 2008). Details about the selection of the experimental population by our partner organization are in the Supplemental Materials available online.⁴

The random assignment procedure placed 143,978 voters in the control group, 133,099 voters in Treatment A, 113,911 voters in Treatment B, 123,077 voters in Treatment C1, and 123,419 voters in Treatment C2.⁵ Table S1 in the Supplemental Materials shows the random assignment process produced treatment groups well balanced on observable covariates, including age, gender, race, past voting record, and state of residence.

Treatments: Varying Call Quality

For the experiment, bids were solicited by our partner organization from multiple firms offering voter mobilization phone calls to civic and political organizations. Our partner organization selected four firms based on expectations of different call delivery quality from three different operating models. Since we exploit the natural

² The states were Florida, Iowa, Illinois, Maine, Michigan, Minnesota, New Mexico, New York, Ohio, Pennsylvania, and South Carolina.

³ Our partner organization is a not-for-profit non-partisan 501(c)3 charitable organization whose mission includes increasing voter participation. The name of the organization is withheld in accordance with our partnership agreement. The partnership agreement specified unrestricted publication rights using the data from this experiment, thus avoiding the potential for selection bias in reported results when organizations control the release of information (Nickerson 2011).

⁴ The results of any experiment are necessarily specific to the context in which they are conducted. Conducting this experiment in partnership with a civic organization makes it more realistic, but means our subjects are not fully representative of all registered voters. That said, the experiment contains a broad cross-section of the electorate across a diverse set of states (see Table S1 in the Supplemental Materials available online).

⁵ Due to the capacity of the call center in treatment A, slightly fewer subjects were assigned. Consequently, more subjects were assigned to the expected high quality calls in Treatment B.

variation in turnout across the operating models for the call centers, each call center was allowed to follow their standard procedures for selecting, training and supervising callers, and managing the list of subjects to call. The script for the calls was held constant across all four call centers.

Our partner organization sought to maximize the impact on mobilization in the 2010 election, as well as testing the effectiveness of the different types of call centers. Therefore, the firms were informed they were part of a competition employing a randomized clinical trial design, with the goal of eliciting each firm's best performance. The awareness of being judged might widen differences between strong and weak firms or it might narrow differences that would occur without it. Nevertheless, each firm included in the study was treated equally by our partner organization and us. Consequently, the competition aspect of the research design does not interfere with the internal validity of the experiment. Any impact of the competition between the firms on the external validity of the experiment depends on how differently these firms perform when they are not being monitored. However, given that organizations paying for calling programs regularly monitor the progress of their contracted call centers, we believe the conditions of our experiment are quite similar to the normal context of voter mobilization programs.

Our partner organization had several criteria for call center selection that address potential concerns about geographic location contributing to differences in effectiveness. First, the call centers had to be located in the United States. Second, individual callers were required to be fluent and accent-free in American English since they would be making calls to eleven states ranging across the Northeast, South, Mid-Atlantic, Midwest, and Mountain West. Our monitoring of the calls at each call center did not identify any callers with noticeable accents. All four calling centers were located in states in the program (one in Minnesota, one in Pennsylvania and two in Florida). Although we were not aware of it when the firms were selected, both treatments C1 and C2 were outsourced to call centers in Florida. Attributing the differences seen below to a "Florida effect" rather than call quality would require rather far-fetched arguments, particularly since this reasoning would also require arguing that the results obtained by the call centers in Minnesota and Pennsylvania should be highly similar.

Table 1 focuses on three aspects of the business practices of the firms selected by our partner organization: in-house versus out-sourced call center, selection of callers, and training of callers. The differences between the firms are clear in the quotes presented from each firm's materials (further description of the firms is in the Supplemental Materials available online).

We expected calls to be high quality when delivered by a call center operated specifically for the purpose of making calls related to politics. This type of call center is relatively rare in the political phone call industry. Treatment A was calls delivered by a firm operating its own call center specifically for voter mobilization and other political calling projects. This firm reports creating an in-house call center to attempt to make volunteer quality calls with "hands on training and monitoring" because the "quality of the phoner is paramount" for effective "chatty conversation" calls. They explicitly contrast their approach to the "large telemarketing centers where speed is critical and quality is less manageable". We expected this

Table 1 Description of call center type, caller selection, and training for each treatment

| | Treatment A | Treatment B | Treatment C1 | Treatment C2 |
|---|--|---|--|---|
| Call center: in-house versus out-source | <p>However, as most groups and campaigns do not have sufficient volunteer resources we have started our own in-house call center... which provide us hands on training and monitoring. We do not prefer large telemarketing centers where speed is critical and quality is less manageable</p> | <p>In addition, our call centers currently have several hundred callers on staff who come to work primarily on the basis of our client list. (Firm B) works only for (<i>redacted description of client list</i>) and we hire our callers around their commitment to these causes and values not because they want to be telemarketers. We want employees who are passionate and articulate as these are the individuals who will be most effective calling on behalf of our clients. (Firm B) is a unionized company which provides health insurance and vacation benefits for full and part-time employees. The average tenure of our staff is 18 months. Many veteran callers have been with us for many years... Because the majority of the work we do is talking with members and donors to secure donations on behalf of our clients we are very comfortable with a conversational vs. a scripted approach</p> | <p>[O]ne of the most important factors before administering a program is ensuring that the call centers we strategically partner with meet our stringent quality standards</p> | <p>[W]e can place the whole thing in one or two call centers to facilitate monitoring and improve quality control</p> |

Table 1 continued

| | Treatment A | Treatment B | Treatment C1 | Treatment C2 |
|------------------|---|---|---|--|
| Caller selection | While the script is certainly an important factor in achieving these conversations, I have found the quality of the phoner is paramount. I will be personally involved in the placement and training for these programs | | [O]ur telephone service representatives understand that in politics we are given one bite at the apple which means those who graduate “phone school” and are selected for political calling are of the upmost quality and capable of delivering complicated scripts with the ease of a two-way, friendly conversation | “All call center employees go through extensive training prior to making any calls. Besides the training necessary to properly utilize the equipment, each caller is trained on the specific script, pronunciations, background information and ultimate goal for the call.” |
| Training | [T]raining them [callers] as to the best way to achieve the “chatty conversation”, that we both know works so well As indicated above, our training procedures are project specific. We conduct early calls in-house to make script adjustments and establish a baseline for training procedures based on both our observations as well as yours | We would begin the program with training around the goals of the program as well as scripting and appropriate data entry. Each manager assigned to the program will make a series of test calls so that they are familiar with the script mechanics and able to better coach calling staff. Each caller assigned to the program will go through script training and post calling we will review and retrain as needed any staff having difficulty. Our quality assurance procedures will include real time monitoring by call center managers along with data review by our Quality Assurance Department to ensure script accuracy and proper data collection | “Before going live with a phone program all of the callers and managers must: 1. Be well versed in the script, knowing all names of candidates, election date, candidate’s party, and basic background information. 2. Understand the goal of the program. 3. Know standard rebuttals and referral information (website or phone number) for voters who have more in-depth questions.” | |

Merged cells indicate the reported quote from the firm covers multiple aspects. All quotes taken from proposals to our partner organization about making the calls. We report only the pre-calling information used in developing expectations of call quality. Further communication during and after the calling confirmed the descriptions in the proposals. See the Supplemental Materials available online for more details

approach to recruitment, supervision, and training of callers in a call center explicitly created for political calling to produce high quality calls.

We also expected high-quality calls from call centers that normally solicit donations for non-profit and political organizations. Treatment B was calls delivered by a firm operating a call center normally specializing in fundraising for political and non-profit organizations.⁶ Although this type of call center may not have extensive experience with voter mobilization calls, they recruit, train, and retain callers to engage in high-quality interactions to successfully solicit contributions for non-profit and political organizations. The firm reported “our call centers currently have several hundred callers on staff who come to work primarily on the basis of our client list... and we hire our callers around their commitment to these causes and values not because they want to be telemarketers. We want employees who are passionate and articulate... Because the majority of the work we do is talking with members and donors to secure donations on behalf of our clients we are very comfortable with a conversational versus a scripted approach”, (see Table 1). Moreover, since these fundraising call centers operate year-round, they are likely to attract and retain the best callers available. “[Firm B] is a unionized company which provides health insurance and vacation benefits for full and part-time employees. The average tenure of our staff is 18 months. Many veteran callers have been with us for many years”. Our partner organization verified this firm’s track record in raising money over the phone. We assumed that skill and experience in influencing one form of civic behavior (contributing to organizations) would translate to influencing another civic behavior (voting).⁷

We expected calls delivered by multi-purpose telemarketing call centers to be of lower quality. The most prevalent operating procedure in the market for professional voter mobilization calls is a political consulting firm that out-sources the call delivery to multi-purpose telemarketing call centers. These political consulting firms offer services to develop scripts, locate call center capacity, train callers, and remotely supervise calls, but the political consulting firms do not have direct control over key aspects of call delivery. Since firms out-sourcing calls to multi-purpose telemarketing call centers are the modal operating procedure in the political calling industry, two of these firms were included in the experiment to assess whether there is variation within this operating model (treatments C1 and C2). We expect lower-quality calls from these firms because the calls are actually made in call centers have little incentive to select, train, or retain people to make calls related to politics since this is a small and seasonal part of their business. These firms’ descriptions of caller selection and training, in Table 1, reveal limited control over the selection of callers and training that is focused on basic competence rather than effective delivery (e.g. “pronunciation”). Previous field experiments relying on calls from these types of call centers have described the delivery of calls as rushed or harried (Nickerson 2007; Gerber and Green 2000). The financial model for these

⁶ Survey research call centers might be a fifth operating model. However, using survey research call centers in this experiment was not possible due to capacity and cost constraints. The higher costs suggest that they might deliver even higher quality calls, although cost is not a guarantee of quality.

⁷ Nickerson (2007) expresses similar expectations about the quality of phone calls made by fundraising call centers.

multi-purpose telemarketing call centers is low profit margin and high volume, so callers are incentivized and pressured to complete calls rapidly.

The motivation for using these four treatments is two-fold. First, they cover the range of voter mobilization phone campaign practices (see Gerring 2001, pp. 218–219, on the “typical case” research design). Second, the four treatments used the same mode of communication (phone calls), the same psychological mechanisms (same script), and were applied at the same time to randomly assigned groups of voters. Thus, different outcomes are caused by the variation in the *delivery* of the calls due to the different call center operating models (see Gerring 2001, pp. 209–212, on the “most similar” research design).

Call Script

All four treatment conditions used an identical phone script that drew upon recent successful voter mobilization field experiments (see Supplemental Materials online for complete script).

- Nickerson and Rogers (2010) found that prompting registered voters to make a plan for voting (implementation intentions) makes voter mobilization phone calls more effective than conventional appeals to civic duty or requests to pledge to vote (Michelson et al. 2009). Our script asked respondents when they planned to vote and how they planned to get to the polls.
- Panagopoulos (2011) found that thanking registered voters for voting in a recent election increased voter turnout by exerting social pressure in a way that did not provoke anger or backlash. Our script thanks registered voters for voting in recent elections.
- Gerber and Rogers (2009) found increases in self-reported intention to vote when the phone calls provided a positive descriptive social norm about voting. Therefore, our script labels the respondent as “the kind of person who cares about your community and who votes”, and suggests joining “the thousands of people like you who will vote on Tuesday”.
- The script also asked respondents to pledge to “fill out the entire ballot” for all candidates and referendums to reduce roll-off for lower salience contests, although we have no expectation that this will influence turnout (Mann 2011).⁸

All four call centers were instructed to make the interactions seem “chatty” and “personal”. We conducted remote telephone monitoring sessions in real time with each firm’s call center throughout the field period. The callers at all of the call centers seemed to have the intended conversational approach, while still adhering to the script. Our experience with this monitoring, coupled with the results below, leads us to recommend more rigorous monitoring in the future. For example, recording phone calls for systematic coding would allow for a more refined assessment of call characteristics.

⁸ The fifth element of the script requested an email address from the respondent to enable our partner organization to send a vote reminder (Dale and Strauss 2007; Malhotra et al. 2011). Less than 1 % of respondents provided emails, so we do not address this part of the script here.

Measurement of Outcomes

We use objective official public records on individual voter turnout to measure call quality via mobilization effects. These data are official public records of individual-level voter turnout in the 2010 General Election, obtained from state and county election administrators by Catalist LLC. Since our original data contained official voter registration numbers, vote validation matching was accurate. These official voting records provide an unbiased measurement of subjects' voter participation.⁹

We use the individual-level turnout data to assess whether the phone calls in each treatment group had the intended effect of increasing turnout. We estimate the mobilization effect by comparing the turnout rate among registered voters assigned to each treatment group to the voter turnout rate among subjects assigned to the control group (average treatment effect, or "ATE"). We also compare the ATEs for each treatment group to one another.¹⁰ Many voter mobilization field experiments focus on the complier average causal effect (CACE) which estimates the increase in turnout *conditional on an indication of treatment delivery*.¹¹ This is inappropriate for our research question because the delivery of the treatment is our quantity of interest. If the treatments reach different proportions or different types of subjects, there will be selection bias in comparing the CACE among the treatments. Therefore, while we report CACE for comparability with other studies, we focus on the ATE for hypothesis testing.

For testing Hypothesis 2—contact rates are a good indicator of call quality—we considered an interaction between the caller and a respondent to be contact [American Association of Public Opinion Research's (2011) Contact Rate 1]. Setting a higher bar for interaction with the subject, we considered a substantive response to the first question in the call script as cooperation with delivery of the treatment [AAPOR (2011) Cooperation Rate 1].¹² Past field experiments using commercial call centers have reported a wide range of contact rates, generally between 30 and 75 % (e.g. see review of 28 such experiments in Gerber and Green (2008, pp. 188–200), but this variation is likely to be caused by differences in jurisdiction, electoral context, target population, and possibly definitions of "contact" as well as differences in call quality. Moreover, there does not appear

⁹ Voters who did not appear on the post-election voter rolls were coded as non-voters. We cannot exclude voters who drop from the voter rolls, because the administrative process for removing a record from the voter rolls is conditional on non-voting under the National Voter Registration Act of 1993. If the treatment increases turnout, it makes voters more likely to remain on the rolls. Thus, exclusion of non-voters from both the treatment and control groups will bias the estimate of the treatment effect.

¹⁰ The ATE (and CACE) estimates use a fixed effects estimator to account for the stratification of the random assignment detailed in the Supplemental Materials online.

¹¹ The CACE reflects the influence of the treatment among those who are successfully treated, (sometimes referred to as a "treatment on treated effect" or "average treatment among the treated effect"). The CACE is the ATE divided by the response rate. In regression analyses, the CACE is estimated using random assignment to the treatment conditions as instruments for successful delivery of the treatment in instrumental variable regression (Gerber and Green 2000; Angrist et al. 1996).

¹² Measuring cooperation with later items in the call script leads to very similar conclusions. Each call center used slightly different codes for call dispositions, so minor measurement error may be introduced by these coding differences.

to be a systematic link between the contact rate and mobilization effect in these past experiments. In short, trying to draw reliable inferences by comparing different experiments is impossible due to unobserved heterogeneity across settings. Thus, the random assignment of different types of call centers in a single experiment provides important causally valid evidence about the relationship between contact rate and mobilization effect.

Results: Voter Turnout

The top of Table 2 reports the ATE on turnout generated by each treatment relative to the control group. Random assignment ensures that these estimates are unbiased. The results in Table 2 are confirmed by regression estimates reported in Tables S2 and S3 in the Supplemental Materials available online.

In line with Hypothesis 1, and our expectations about call quality from each treatment, treatments A and B significantly increased turnout while treatments C1 and C2 generate no increase in turnout. Turnout in the control group was 38.73 %. The ATE on turnout for Treatment A (39.14 %, +0.42 percentage points, $p = 0.013$) and Treatment B (39.43 %, +0.70 percentage points, $p < 0.001$) are statistically significant. Treatments C1 and C2 failed to generate statistically significant increases in turnout at the conventional 95 % confidence level, despite the considerable statistical power of this experiment.

We test Hypothesis 1 about the differences in call quality with pair-wise comparisons: Treatment B generates significantly higher turnout than treatments C1 and C2 (both comparisons $p < 0.01$, one-tailed); Treatment A also causes higher turnout than treatments C1 and C2 but both differences are only marginally statistically significant ($p = 0.075$, one-tailed); Treatment A is statistically indistinguishable from treatment B ($p = 0.16$); and treatments C1 and C2 are also statistically indistinguishable ($p = 0.99$).

While unobserved heterogeneity between experimental settings makes direct comparisons of individual experiments problematic,¹³ it is informative to compare these results to previous experiments. Green and Gerber (2008) performed a meta-analysis with random effects covering 28 field experiments on voter mobilization calls using commercial phone banks. They use the CACE to account for differences in contact rates across experiments. The lower portion of Table 2 presents the contact rate and CACE from our experiment to facilitate this comparison (see Table S3 in the Supplemental Materials for full CACE results). Gerber and Green found a

¹³ Among many potential sources of heterogeneity, voter mobilization treatments are more likely to be attenuated by the mobilization efforts of other groups in the 2010 mid-term general election than in the low salience elections in which many field experiments are conducted. We also refrain from using cost per net vote as a basis of comparison since the calls in the experiment received discounts that bias this comparison based on the number of calls and our partner organization's status as a major (potential) client.

Table 2 Voter turnout in 2010 general election ATE and CACE

| | Control group | Treatment A | Treatment B | Treatment C1 | Treatment C2 |
|--------------------------------|---------------|--------------------------|---------------------------|--------------|--------------|
| Turnout | 38.73 % | 39.15 % | 39.43 % | 38.87 % | 38.87 % |
| Average treatment effect | | 0.42 ⁺ (0.19) | 0.70 ⁺⁺ (0.19) | 0.14 (0.19) | 0.14 (0.19) |
| Contact rate | | 48.05 % | 47.89 % | 75.58 % | 71.25 % |
| Complier average causal effect | | 0.87 ⁺ (0.39) | 1.44 ⁺⁺ (0.41) | 0.18 (0.25) | 0.20 (0.27) |
| <i>N</i> | 143,978 | 133,099 | 113,911 | 123,077 | 123,419 |

Results based on regression estimates without covariates. The models include indicator variables to account for strata in random assignment procedure. Regression estimates with and without full set of covariates are provided in Tables S2 and S3

⁺ $p < 0.05$; ⁺⁺ $p < 0.01$ (one-tailed). Standard errors reported in parentheses

0.55 percentage point weighted mean CACE [confidence interval 0.073–1.034].¹⁴ The CACE from high-quality calls in treatment A (0.87 percentage points) and treatment B (1.44 percentage points) compare favorably to the meta-analysis despite being conducted in a higher salience mid-term general election. In comparison to previous studies, the failure of treatments C1 and C2 to generate any increase in turnout makes them very poor performers, particularly since treatments A and B confirm that the call script is effective.

Results: Contact Rate

To examine Hypothesis 2, Table 3 reports the contact rate and cooperation rate. Table 3 shows a marked difference in the contact rate and cooperation rate across the four treatments, with significantly higher rates for treatments C1 and C2. The contact rates for treatments A and B are both 48 %. Contact rates for treatments C1 (75.6 %) and C2 (71.3 %) are significantly higher. The rate of response to the call script's queries (cooperation rate) has a larger difference, so the differences are not an artifact of "contacts" that did not deliver the treatment. For the cooperation rate, treatments A (14.8 %) and B (18.4 %) are much lower than the 50 % cooperation rate reported for treatments C1 and C2.

The inverse pattern in effects on turnout and contact rate fails to provide any support for the conventional assumption that higher response rates are an indication of high quality voter mobilization calls (Hypothesis 2). More specifically, assuming that higher contact rates indicate better calls would have led to erroneous conclusions about the performance of the respective call centers in mobilizing voters in this experiment. While low contact rates are not desirable since this means fewer treatments are being delivered, increasing contact rates at the expense of quality appears to be a worse alternative for both net effect on democratic

¹⁴ Using a slightly different set of 10 field experiments on voter mobilization calls from commercial call centers, Ha and Karlan's (2009) meta-analysis calculates a 0.9 percentage point weighted mean CACE. Substitution of this meta-analysis estimate for the Green and Gerber (2008) meta-analysis leads to similar substantive conclusions.

Table 3 Contact rate and cooperation rate by treatment

| | <i>N</i> (assigned) | Contact rate (%) | Cooperation rate (%) |
|--------------|---------------------|------------------|----------------------|
| Treatment A | 133,099 | 48.05 | 14.77 |
| Treatment B | 113,911 | 47.89 | 18.43 |
| Treatment C1 | 123,077 | 75.58 | 50.10 |
| Treatment C2 | 123,419 | 71.25 | 49.85 |

Contact rate follows AAPOR's (2011) Contact Rate 1 definition of any interaction of the interviewer and a respondent. Cooperation rate reflects a higher bar of response to the queries in the call script [AAPOR (2011), Cooperation Rate 1]

participation and statistical power to test the content of phone scripts. Moreover, the contrast between the contact rates and the effects on turnout provides further support for call quality as the key factor in voter mobilization. In order to generate larger treatment effects while contacting fewer people, the calls in treatments A and B must have delivered higher quality mobilization stimulus.

Discussion

To assess the effect of call quality on voter mobilization we utilized a large-scale randomized field experiment involving 637,484 registered voters in eleven states in the 2010 General Election. Our standard for call quality is credible and sincere delivery of the voter mobilization message. The experiment exploited natural variation in operating models across call centers to vary call quality for testing whether call quality influences effectiveness at voter mobilization. We found that treatments A and B, which were expected to deliver high quality calls because the call centers operate specifically for calls related to politics and fundraising, generate a statistically significant and substantively large positive effect on voter turnout. Treatments C1 and C2, which were expected to deliver low quality calls because calls were out-sourced to multi-purpose telemarketing call centers, do not increase turnout. Moreover, the effect on turnout from the quality calls delivered by treatments A and B was significantly larger even though treatments C1 and C2 reported contacting a larger portion of their respective treatment groups.

These findings make a number of contributions to our understanding of voting behavior, and how participation in elections can be increased. First, we demonstrate that the quality of treatment delivery influences the efficacy of voter mobilization treatments. Second, these results provide additional insights into the mechanism that makes voter mobilization by strangers effective at overcoming the collective action dilemma of voting: mobilization calls from strangers must deliver a high-quality treatment. Without a high quality—credible and sincere—delivery, otherwise effective treatments will generate no boost in political participation. Third, the results are a reminder that users of field experiments must pay attention to the quality of treatment delivery before drawing substantive conclusions. This is particularly important when experimental evidence fails to reject the null hypothesis of no treatment effects since our results demonstrate that poor treatment delivery

can undermine the observation of a true treatment effect. Fourth, these contributions are practical lessons for organizations seeking to increase voter turnout: work with call centers directly operated by firms engaged full-time in phone calls related to politics.

Contrary to the conventional wisdom, we also found that contact rate may be a misleading measure of quality. Our results show that “less is more”: treatments A and B reported far lower contact rates than treatments C1 and C2, but generate a significantly larger increase in turnout. Contact rates may be a convenient statistic, but the evidence here indicates they are not a useful measure of quality, and may even be a misleading one. Our outcome measure of call quality—the mobilization effect of the call—is a more reliable and objective measure of call quality.

For scholars, the finding that contact rate appears negatively related to overall effect on turnout has an important methodological implication for experimental design. In this experiment, maximizing the contact rate would undermine the statistical power of any test of different content in these phone calls. When seeking to maximize statistical power, focusing on quality of treatment delivery is a more important consideration than the proportion of treatment delivery (i.e. contact rate), because there is a risk that increasing contact rates will undermine quality as seen in this experiment.

While our findings make important contributions to our understanding of voter turnout, our ability to draw conclusions about the micro-level mechanisms underlying these results, such as the characteristics of individual callers, is limited by our reliance on aggregate differences between call centers with different operating models. Thus, ours is a first step that leads to future research focusing more specifically on caller characteristics and other factors that influence call quality. Drawing from the literature on survey research house effects (for a summary, see [Leahey 2008](#) and [Weisberg 2005](#), pp. 299–303), factors that influence the quality and effectiveness of voter mobilization calls could include recruitment and retention of callers, training, monitoring, pay and incentives as well as demographic and attitudinal characteristics of callers.

For example, one possibility we have suggested is that call centers specializing in political calling projects, like treatments A and B, attract and retain callers with a higher level of interest in politics. These callers may transmit their interest in politics with a more enthusiastic tone of voice or other vocal cues without altering the words of the script. Another possibility is that specialized call centers have a more stable set of employees, leading to greater professionalism and relevant skills that translate to more effective stimulation of voter turnout. These specialized call centers may also attract a set of callers with a different demographic profile. Respondents’ perceptions of sex, race, education level, and other characteristics of callers can influence the interviewer-respondent interaction ([Weisberg 2005](#), pp. 58–63). Future field experiments should randomly assign subjects to different types of callers to measure who responds to whom when encouraged to participate in voting.

Low voter turnout is a threat to popular sovereignty, and a symptom of weak democratic institutions. In order for civic and political organizations to successfully increase participation in democratic elections, we must understand how to motivate

citizens to cast a ballot. Our results extend understanding of the influence of call quality on these types of voter mobilization efforts, and lead to new questions on how to increase the efficacy of these efforts. Beyond the narrow, though normatively important, application to voter mobilization tactics, these findings provide useful insights for other fields engaged in contacting citizens and encouraging action, ranging from private sector marketing to public policy advocacy, public health, public safety, and many other public goods.

Acknowledgments We thank our partner organization for the support that made this research possible. We thank the editors, anonymous reviewers, Stephen Ansolabehere, Donald Green, John Love, Frank Sansom, Brian Shaffner, and the participants in the Political Science Faculty Colloquium at the University of Miami for their helpful comments. An earlier version of this paper was presented at the 2012 Midwest Political Science Association Conference. This research was conducted under University of Miami Human Subjects Research Office Protocol #20110124. Replication data is available at <http://www.sites.google.com/site/christopherbmann>. All errors are the responsibility of the authors.

References

- American Association for Public Opinion Research (AAPOR). (2011). *Standard definitions: Final dispositions of case codes and outcome rates for surveys* (7th ed.). Washington: AAPOR.
- Angrist, Joshua D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, *91*(434), 444–455.
- Atkeson, L., Adams, A., Bryant, L., Zilberman, L., & Saunders, K. (2011). Considering mixed mode surveys for questions in political behavior: Using the internet and mail to get quality data at reasonable costs. *Political Behavior*, *33*, 161–178.
- Bedolla, L. G., & Michelson, M. R. (2012). *Mobilizing inclusion: Transforming the electorate through get-out-the-vote campaigns*. New Haven: Yale University Press.
- Dahl, R. A. (1956). *A preface to democratic theory*. Chicago: University of Chicago Press.
- Dale, A., & Strauss, A. (2007). Don't forget to vote: Text message reminders as a mobilization tool. *American Journal of Political Science*, *53*(4), 787–804.
- Gerber, A. S., & Green, D. P. (2000). The effects of personal canvassing, telephone calls, and direct mail on voter turnout: A field experiment. *American Political Science Review*, *94*, 653–664.
- Gerber, A. S., & Green, D. P. (2001). Do phone calls increase voter turnout? A field experiment. *Public Opinion Quarterly*, *65*(1), 75–85.
- Gerber, A. S., & Green, D. P. (2005a). Correction to Gerber and Green (2000), replication of disputed findings, and reply to Imai. *American Political Science Review*, *99*(2), 301–313.
- Gerber, A. S., & Green, D. P. (2005b). Do phone calls increase voter turnout? An update. *Annals of the American Academy of Political and Social Science*, *601*(1), 142–154.
- Gerber, A. S., & Rogers, T. (2009). Descriptive social norms and motivation to vote: Everybody's voting and so should you. *Journal of Politics*, *71*(1), 178–191.
- Gerring, John. (2001). *Social science methodology: A criterial framework*. New York: Cambridge University Press.
- Green, Donald P., & Gerber, A. S. (2008). *Get out the vote!* (2nd ed.). Washington: Brookings.
- Ha, S. E., & Karlan, D. S. (2009). Get-out-the-vote phone calls: Does quality matter? *American Politics Research*, *37*(2), 353–369.
- Keeter, S., Kennedy, C., Dimock, M., Best, J., & Craighill, P. (2006). Gauging the impact of growing nonresponse on estimates from a national RDD telephone survey. *Public Opinion Quarterly*, *70*(5), 759–779.
- Klofstad, C. A., McDermott, R., & Hatemi, P. K. (2011). Do bedroom eyes wear political glasses? The role of politics in human mate attraction. *Evolution and Human Behavior*, *33*(2), 100–108.
- Leahey, E. (2008). Methodological memes and mores: Toward a sociology of social research. *Annual Review of Sociology*, *34*, 33–53.

- Malhotra, N., Michelson, M. R., Rogers, T., & Valenzuela, A. A. (2011). Text messages as mobilization tools: The conditional effect of habitual voting and election salience. *American Politics Research*, 39(4), 664–681.
- Mann, C. B. (2011). *Preventing ballot roll-off: A multi-state field experiment addressing an overlooked deficit in voting participation*. New Orleans: Annual Meeting of the Southern Political Science Association.
- Michelson, M. R., Bedolla, L. G., & McConnell, M. A. (2009). Heeding the call: The effect of targeted two-round phone banks on voter turnout. *Journal of Politics*, 71(4), 1549–1563.
- Nickerson, D. W. (2006). Volunteer phone calls can increase turnout: Evidence from eight field experiments. *American Politics Research*, 34(3), 271–292.
- Nickerson, D. W. (2007). Quality is job one: Professional and volunteer voter mobilization calls. *American Journal of Political Science*, 51(2), 269–282.
- Nickerson, D. W. (2008). Is voting contagious? Evidence from two field experiments. *American Political Science Review*, 102(1), 49–57.
- Nickerson, D. W. (2011). When the client owns the data. *The experimental political scientist: Newsletter of the APSA experimental section*, 2(2), 5–6.
- Nickerson, D. W., & Rogers, T. (2010). Do you have a voting plan? Implementation intentions, voter turnout, and organic plan making. *Psychological Science*, 21(2), 194–199.
- Olson, M. (1965). *The logic of collective action: Public goods and the theory of groups*. Cambridge: Harvard University Press.
- Olson, K., & Bilgen, I. (2011). The role of interviewer experience on acquiescence. *Public Opinion Quarterly*, 75(1), 99–114.
- Panagopoulos, Costas. (2011). Thank you for voting: Gratitude expression and voter mobilization. *Journal of Politics*, 73(3), 707–717.
- Porter, T. (1995). *Trust in numbers*. Princeton: Princeton Univ. Press.
- Sinclair, B., McConnell, M. A., & Green, D. P. (2012). Detecting spillover effects: Design and analysis of multi-level experiments. *American Journal of Political Science*, 5(4), 1055–1069.
- Smith, T. W. (1978). In search of house effects: A comparison of responses to various questions by different survey organizations. *Public Opinion Quarterly*, 42(4), 443–463.
- Smith, T. W. (1982). House effects and the reproducibility of survey measurements: A comparison of the 1980 GSS and the 1980 American national election study. *Public Opinion Quarterly*, 46, 54–68.
- Viterna, J. S., & Maynard, D. W. (2002). How uniform is standardization? Variation within and across survey research centers regarding protocols for interviewing. In D. W. Maynard, H. Houtkoop-Steenstra, N. C. Schaeffer, & J. van der Zouwen (Eds.), *Standardization and tacit knowledge: Interaction and practice in the survey interview*. New York: Wiley.
- Weisberg, H. F. (2005). *The total survey error approach*. Chicago: University of Chicago Press.