RUNNING HEAD: Outsized turnout effects of subtle linguistic cues

Do Subtle Linguistic Interventions Priming a Social Identity as a Voter Have Outsized Effects on Voter Turnout? Evidence from a New Replication Experiment

May 12, 2017

ABSTRACT: An ongoing debate in political psychology is about whether small wording differences have outsized behavioral effects. A leading example is whether subtle linguistic cues embedded in voter mobilization messages dramatically increase turnout. An initial study analyzing two small-scale field experiments argued that describing someone as a voter (noun) instead of one who votes (verb) increases turnout rates 11 to 14 points because the noun activates a person's social identity as a voter. A subsequent study analyzing a large-scale field experiment challenged this claim and found no effect. But questions about the initial claim's domain of applicability persist. The subsequent study may not have reproduced the conditions necessary for the psychological phenomenon to occur, specifically the electoral contexts were not competitive or important enough for the social identity to matter. To address the first of these critiques, as well as other potential explanations for different results between the first two studies, we conduct a large-scale replication field experiment. We find no evidence that this minor wording change increases turnout levels. This research provides new evidence that the strategy of invoking the self does not appear to consistently increase turnout and calls into question whether subtle linguistic cues have outsized behavioral effects.

KEYWORDS: political psychology; intervention; field experiment; voter turnout; participation; linguistic cues

ABSTRACT WORD COUNT: 200
MANUSCRIPT WORD COUNT: 6238

An ongoing debate in political psychology is about whether small wording differences have outsized effects on behavior. A leading example is whether subtle linguistic cues that link a desired behavior to a person's social identity are able to induce dramatic behavioral change by priming that identity. Specifically, the argument is that describing a person using a predicate noun (e.g., "to be a voter") emphasizes a behavior as an attribute of one's social identity that one can claim by engaging in that behavior. In contrast to describing a person's potential behavior using a verb (e.g., "to vote"), the theory argues that the use of a predicate noun introduces a subtle linguistic cue that more clearly primes the behavior as related to one's identity, increasing the likelihood that the person engages in it. Past research has sought to show how noun wording affects both contemporaneous non-political behavior (Gelman and Heyman, 1999; Walton and Banaji, 2004) as well as prospective political behavior (Bryan et al., 2011).

As part of a broader class of so-called "wise interventions" (Walton, 2014), these studies are important for both theoretical and policy-relevant reasons. First, they represent a constructive effort among researchers to explicitly theorize the psychological processes that explain why priming interventions might plausibly cause changes in political behavior. Accordingly, empirically testing the effectiveness of these priming interventions is important to assess the validity of psychological theories of behavioral change that are both ubiquitous and understudied in the political participation literature. Second, early studies reporting large effects have been influential in shaping the design of real-world interventions and how campaign professionals have chosen to allocate scarce resources to affect political behavior and electoral outcomes. These studies represent a broader class of priming experiments where minor linguistic interventions are viewed as novel and promising because they are perceived to have major behavioral implications, to be low cost, and to be portable to many choice environments. Most generally, the perception that minor linguistic interventions generate high returns on investment has contributed to their widespread adoption.

Despite their *prima facie* appeal and early published reports of large effects, there remains a need to understand whether and why these interventions are effective at increasing political participation and whether initially reported large effects are reproducible and robust. We argue that

making scientific progress to answer these questions requires two elements. First, for a specific intervention designed to affect a specific behavior, multiple, well-powered replication experiments are necessary (i) to assemble a sampling distribution of the effect of that intervention on the behavior of interest and (ii) to rule out the possibility that observed and published results are driven by sampling variability and sample selection. Second, in the event replications generate compelling evidence of null effects, researchers should revisit whether the intervention tested is, in fact, able to affect the psychological process of interest. In this article, we address both of these needs as they apply to the study of subtle linguistic cues that use predicate nouns to encourage political participation by priming a citizen's social identity as a voter.

In an influential article published in the *Proceedings of the National Academy of Sciences*, Bryan et al. (2011) explore the effect of psychologically inspired interventions in political settings and test the validity of this hypothesis among adults in a political setting using a series of three experiments that assess the effects of noun wording instead of verb wording on future political participation.[1] They find a large effect of the noun wording condition, relative to the verb condition, on stated interest in political participation (Study 1) and on observed turnout levels in two field experiments (Studies 2 and 3). Subjects who answer 10 questions on an Internet survey that use the noun wording (referring to the subject as a voter) instead of 10 parallel questions that use the verb wording (referring to the subject as a person who votes) express higher levels of interest in voting, and registered voters in California and New Jersey are more likely to vote by 11 to 14 percentage points. These reported effects are substantively and comparatively large in magnitude. Prior field experimental research evaluating the effectiveness of direct mail and phone-based voter mobilization strategies found, for instance, that the most effective appeals to vote are those that invoke social pressure and that they increase turnout by 2.3 and 2.9 points when the appeals are delivered by mail and by phone, respectively (Green and Gerber, 2015).

In light of the surprising magnitude of the effects reported by Bryan et al. (2011), other scholars have sought to replicate these findings and assess their robustness. During the 2014 midterm

[1] This article has been cited 100+ times according to Google Scholar (retrieved January 30, 2017).

primary elections, Gerber et al. (2016) conducted a placebo-controlled field experiment that replicated and extended the Bryan et al. (2011) study with more than 20 times the number of subjects from the original Bryan et al. (2011) field experiment. Their sample was comprised of registered voters in three states (Michigan, Missouri, and Tennessee) who had valid voter file records and who could be contacted by phone. They find no difference between the turnout rates of subjects assigned to the noun versus verb wording conditions, and that both noun and verb treatments are less effective at increasing turnout than a standard "information only" Get-Out-The-Vote (GOTV) treatment message. These findings cast doubt on the robustness of the results by Bryan et al. (2011) while also raising questions about the conditions under which the original results are reproducible.

However, multiple study parameters vary across these studies, including the mode of treatment delivery [via Internet survey for Bryan et al. (2011) versus phone for Gerber et al. (2016)], the electoral context [Gerber et al. (2016) focus on 2014 House primary elections in 3 states whereas Bryan et al. (2011) focus on the 2008 general election in California and the 2009 New Jersey gubernatorial election], subjects' demographic characteristics [Gerber et al. (2016) employ a substantially more diverse sample than Bryan et al. (2011)], and other factors affecting selection into the subject pool. In a response to Gerber et al. (2016), three of the four original authors from Bryan et al. (2011) state it is important for replication experiments to reproduce "the psychological context in which the phenomenon could plausibly emerge" by replicating the procedures of their original study in comparable electoral contexts (Bryan et al., 2016). Specifically, the primary criticism made by Bryan et al. (2016) is that the psychological context was not reproduced because the election contexts from the Gerber et al. (2016) study were not sufficiently competitive or important for a person's identity as a "voter" to be salient enough to motivate behavioral change.[2]

[2]Bryan et al. (2016) write in their response to Gerber et al. (2016):

> "We tested our hypothesis in high-profile elections that received substantial public attention. [...] In these cases, the opportunity to be 'a voter' feels like a valued identity that can motivate behavior. [...] Contrast this with the congressional primaries examined in Gerber et al. (1): The outcomes of almost none were in doubt—nearly half were uncontested—and few received any meaningful attention. Consider the 12 of 61 major party House primaries that Gerber et al. call "competitive": [...] one was uncontested and six others were decided by huge margins (20.5, 32.8, 33, 39, 47.8, and 58.6 percentage points). In reality, only 4 of those 61 primaries were meaningfully competitive by even a loose standard. [...] Psychological experiments... are predicated on a careful analysis of psychological processes and the contextual factors that influence them. An authentic replication begins with this psychological

This article describes and presents the results from a new experiment that contributes to this ongoing debate. We provide additional evidence to assess the hypotheses that political participation can be increased by cueing one's identity as a voter through the use of predicate noun language (instead of verb wording). Two core features of this experiment directly address concerns about how the Gerber et al. (2016) study was different from the studies reported in Bryan et al. (2011).

First, the replication experiment was conducted in 2015 in the context of contested gubernatorial elections in Kentucky, Louisiana, and Mississippi and of the contested mayoral election in Houston, Texas. These contexts were selected to address the primary criticism made by Bryan et al. (2016) about the Gerber et al. (2016) study. In particular, the contests in Kentucky, Louisiana, and Houston were all considered to be competitive both *ex ante* and *ex post*, and thus plausibly "allow the identity 'voter' to feel important" and thereby motivate behavior (Bryan et al., 2016). With respect to *ex ante* competitiveness, the gubernatorial elections in Kentucky and Louisiana were both contests over open seats and were both rated as toss-up races by Cook Political Report.[3] The mayoral election in Houston was a competitive six-way race, where no candidate was expected to win a majority and where the top two candidates would face off in a subsequent runoff election.[4] To assess *ex post* competitiveness, we examine the actual vote margins (i.e., the difference in vote shares between the winning candidate and the runner-up) in each of these races. As Table 1 shows, the actual winning vote margins are generally smaller (as compared to those examined in prior studies) and suggest meaningfully competitive races: 8.7 points in Kentucky (52.5% to 43.8%), 6 points between the top two candidates and 8.1 points between the second and third place candidates in the Houston mayoral election (31.3% to 25.2% and 25.2% to 17.1%, respectively), and 12.2 points in Louisiana (56.1% to 43.9%). By contrast, for the electoral contexts from the two field experiments in Bryan et al. (2011), the actual vote margins in percentage points were 24.1

understanding (3). In the best cases it attempts to extend this understanding by, for example, directly comparing contexts in which a phenomenon is likely to emerge to ones in which it is not. In the present case, an electoral context that allows the identity 'voter' to feel important is necessary to motivate behavior." (E6548).

[3] `http://web.archive.org/web/20151107154359/http://cookpolitical.com/governor/charts/race-ratings`.

[4] `https://www.texastribune.org/2015/10/27/houston-mayoral-race/`.

points in the 2008 presidential election in California (we focus on the vote margin in California instead of the national popular vote margin because in presidential general elections, the Electoral College awards votes to candidates by state) and 3.6 points in the 2009 gubernatorial election in New Jersey. In the congressional primary elections in Gerber et al. (2016), the mean winning vote margin across district-primary election contexts were larger and ranged from 22.9 to 67 points.

[Table 1 about here]

Second, the experiment replicates the mode of treatment delivery in Bryan et al. (2011) in an attempt to reproduce the psychological context experienced by subjects in the original study. Specifically, the treatment scripts tested in this experiment (which are identical to those used in the Bryan et al. (2011)) are delivered via an Internet survey, which is the procedure used in Bryan et al. (2011). Finally, we conduct this replication experiment using two different survey firms. One pre-matches their respondents to state voter files and the other requires post-matching respondents to voter files. We do this to address the potential criticism that differences in findings from prior studies may result from differences in sample selection and measurement procedures. Thus, in addition to addressing the main criticism articulated by Bryan et al. (2016) about reproducing a psychologically relevant and electorally competitive context, our replication experiment addresses other potential explanations for the differences in results reported by Bryan et al. (2011) and Gerber et al. (2016) that relate to differences in the design and implementation of the experiments.

Analyzing data from this new replication experiment, we find no evidence that the noun wording treatment increases turnout more than the verb treatment. This result is consistent across electoral contexts and vendors used to deliver the treatment. The 95% confidence interval of the estimated noun wording effect from our primary and preferred specification is (-0.017, 0.031), which excludes the 11 to 14 percentage point effects reported in the initial study by Bryan et al. (2011).[5] These findings therefore provide compelling evidence that priming a subject's identity as a voter using noun wording in an online survey has no effect on turnout when compared to a counter-

[5]Similarly, the 95% confidence interval for the largest subgroup-specific estimate of 0.03 (for the SSI subsample) is (-0.008, 0.069), which also excludes the prior effects reported by Bryan et al. (2011).

factual online survey that employs verb wording. Our findings suggest that the minor use of a predicate noun (e.g., "to be a voter") instead of verbs (e.g., "to vote/voting") in 10 question items presented in an online survey does not affect voting behavior, even if the cue is delivered in the two days leading up to and including Election Day in high-salience and competitive electoral contexts. Importantly, we stress that our results do not suggest that all interventions that attempt to induce behavioral change by priming social identities are categorically ineffective. Rather, our results suggest that some of the possible explanations for why the ineffectiveness of the minor linguistic intervention initially evaluated by Bryan et al. (2011) do not appear to account for the consistent null results that exist in a growing body of subsequent evidence.

The remainder of the article proceeds as follows. In Section 1 we briefly summarize the design details and results from prior studies by Bryan et al. (2011) and Gerber et al. (2016). Section 2 describes the details of our new experiment and analysis plan. We report results from our study in Section 3 and discuss our findings in Section 4.

1 PRIOR STUDIES

We briefly describe the design and findings from prior studies in this section before presenting the details of our replication study. Table 2 presents a summary of prior studies.

[Table 2 about here]

The original set of studies examining the effects of noun wording on turnout by Bryan et al. (2011) consists of three experiments. The first experiment, which was embedded in an online survey fielded during the 2008 presidential election, included 34 subjects who were eligible voters in California, not yet registered to vote, and native English speakers. Subjects were randomly assigned to receive a 10-item questionnaire asking about one's thoughts about being a voter (noun condition) or a different 10-item questionnaire asking about one's thoughts about voting (verb condition) in the upcoming election. The only difference between the questions across treatment arms is whether the question wording uses a predicate noun or a verb (e.g., "How impoortant is it to

you to (vote/be a voter) in the upcoming election?"). Each question had a five-point response scale with category labels specific to the content of each question (e.g., "Not at all important, Not too important, Neither important nor unimportant, Somewhat important, or Extremely important").[6] Subjects were also asked, after the noun or verb questionnaire, whether they were more interested in registering to vote and supplied their response using a 5-point scale measured from "not at all interested" to "extremely interested." On this particular item, compared to an average response of 3.4 among the verb wording group, the noun wording group reported a higher level of interest in registering to vote with an average response of 4.4.

The second and third experiments conducted by Bryan et al. (2011) assess the effect of noun versus verb wording on turnout in the November 2008 and November 2011 general elections, respectively. Subjects in these two studies are drawn from an online survey subject pool and thus consist of individuals who regularly participate in survey experiments. The subject pool was further restricted to native English speakers, those who self-report being registered to vote, and those who self-report as having not yet voted at the time of the survey in California and New Jersey, respectively.[7] Bryan et al. (2011) report that the noun wording increases turnout levels by 13.7 points when compared to a verb condition turnout rate of 81.8% in Experiment 2 and by 10.9 points when compared to a verb condition turnout rate of 79% in Experiment 3.

Gerber et al. (2016) conducted an experiment during the 2014 primary elections in Michigan, Missouri, and Tennessee that employs the treatments used in Experiment 3 from Bryan et al. (2011) and extends it by adding a placebo condition and an "information-only" GOTV condition to assess the comparative effectiveness of noun- and verb-based cueing on turnout.[8] Unlike in Bryan et al. (2011), the treatments in this replication study were delivered by telephone in the 4 days leading up to and including Election Day, and subjects were defined as registered voters in these three states

[6]Online Supplemental Appendix A contains the complete text of the treatment scripts used in all 3 experiments in Bryan et al. (2011).

[7]The question wording in these two experiments deviate from the question wording in the first experiment in two minor ways. First, the question wording for the fifth question is simplified while preserving its meaning. Second, whereas questions in the first experiment refer to the election as "the upcoming election," the second experiment always refers to "tomorrow's" election and the third experiment refers to either "tomorrow's" or "today's" election depending on the day the survey experiment is fielded.

[8]Online Supplemental Appendix B contains the complete text of the treatment scripts used in Gerber et al. (2016).

with valid voter file records and who could be contacted by phone. This experiment included more than 20 times the number of subjects from the larger Bryan et al. (2011) study assessing effects on turnout (4468 subjects assigned to the noun or verb conditions versus 214 subjects in Experiment 3 by Bryan et al.), which provides ample statistical power to detect effects should they exist.[9] Gerber et al. (2016) find that the effect of noun wording statistically indistinguishable from zero and infer that there is no evidence of a difference in turnout rates between treatment arms.[10]

The results of Bryan et al. (2011) and Gerber et al. (2016) are very different, but as previously explained there were many differences in the design and context across these studies as well. We therefore conducted a replication study that addresses some of these differences, chief among them the primary criticism made by Bryan et al. (2016) that the Gerber et al. (2016) study did not occur in the context of an election that was competitive enough.

2 STUDY DETAILS: DESIGN, IMPLEMENTATION, AND ANALYSIS

Our field experiment was conducted during the 2015 general election in three states with a gubernatorial election (Kentucky, Louisiana, and Missisippi) and in a large city with a contested mayoral election (Houston, TX).

Subject recruitment and sample definition. Subjects were recruited from two panels, one administered by Survey Sampling International (SSI) and one by YouGov. Both SSI and YouGov recruited respondents from existing online survey panels, targeting individuals who lived in the selected geographies. Respondents from both online panels were recruited to take a survey in the four days leading up to and including Election Day. In Kentucky, Mississippi, and Houston,

[9] We conduct *post hoc* power analyses and find that Study 2 by Bryan et al. (2011) is designed to detect a 17.3 percentage point effect at 80% power (they report a 13.7 percentage point effect) and that Study 3 by Bryan et al. (2011) is designed to detect a 13.4 percentage point effect (they report a 10.9 percentage point effect). To achieve 80% power, Bryan et al. (2011) would have needed 161 subjects (instead of 88 subjects) in Study 2 and 340 subjects (instead of 214 subjects) in Study 3.

[10] Gerber et al. (2016) report that this finding is robust to the use of non-parametric versus parametric estimators; to partitioning the data by state, district electoral context, or past patterns of voter participation; to stricter definitions of treatment by conditioning on completing all 10 survey items; and to partitioning on subjects contacted the two days leading up to and including Election Day (to replicate the timing of treatment delivery in Bryan et al. (2011)). While these robustness checks do not estimate noun effects on random partitions of the data, they provide observational evidence suggesting that the null effect is robust.

the survey experiment was implemented between October 31, 2015, and the general election on November 3, 2015. In Louisiana, the survey experiment was implemented between November 18, 2015, and the runoff election on November 21, 2015. Within each target area, individuals recruited to participate in the study were randomly sampled to be recruited either 3 days prior to Election Day with 25% probability, 2 days prior with 25% probability, or either 1 day prior to or on Election Day with 50% probability. We stopped recruiting subjects at the time the polls closed on Election Day in each jurisdiction.

Respondents recruited for the SSI sample were admitted into the experiment if their reported 5-digit zip code for their place of residence is contained in a master set of zip codes associated with Kentucky, Louisiana, Mississippi, or Houston[11]; if they reported being registered to vote at the time of the survey; if they are at least 18 years old, which is derived from their reported birth year; and if they provided a valid first and last name.[12] Importantly, these screeners were implemented prior to randomization and therefore selection into the sample has no effect on the internal validity of our noun wording effect estimates.

Respondents recruited for the YouGov sample were admitted into the experiment if they were identified by YouGov as a registered voter age 18 or older and if they reported living in a zip code contained in master zip code list.[13] There was no need to ask subjects in the YouGov sample about their voter registration status, year of birth, or name because YouGov had previously collected the

[11]To produce the master zip code list, we obtained one geographic equivalency file mapping 5-digit ZIP/ZIP Census Tabulation Areas (ZCTA5) in 2010 to states for Kentucky, Louisiana, and Mississippi and another equivalency file mapping ZIP/ZCTA5s to Houston (using the Houston Place FIPS code 35000) from the Missouri Census Data Center's MABLE/Geocorr12 Geographic Correspondence Engine (`http://mcdc.missouri.edu/websas/geocorr12.html`). For the ZIP/ZCTA5-to-state crosswalk file, we retained all ZIP/ZCTA5s that were fully contained in a target state and ZIP/ZCTA5s for which at least 90% of the ZIP/ZCTA5-level population (per the 2010 Decennial Census) was included in one of the target states. For the ZIP/ZCTA5-to-Houston crosswalk file, we retained all ZIP/ZCTA5s that were fully contained in Houston or for which at least 80% of the ZIP/ZCTA5-level population (per the 2010 Decennial Census) was included in Houston.

[12]We ask all respondents for their first and last name and consider the responses provided as valid only if they provide at least two characters for each field. We do this in order to screen out individuals who do not provide adequate information about their full name, which is needed to match survey respondents to state voter file records. A potential concern and source of attrition arises if respondents provide more than two characters in each of the name fields but provide false names. We do not believe this is a major concern, however. Of the 1788 subjects in the SSI sample, 49 (2.7%) had problematic names: 48 did not provide a valid name per our screening criteria (i.e., at least one of the names provided contained less than 2 letters) and 1 wrote "none" for their first and last name.

[13]The master zip code list is the same as the one used for the SSI sample.

personal identifying information necessary to match subjects to state voter files. In total, 3608 respondents were admitted to the experiment as subjects, of whom 1848 are recruited from the YouGov panel and 1230 are recruited from the SSI panel.

Treatment scripts. Respondents admitted into the experiment as a subject are immediately randomly assigned to receive either a 10-item questionnaire using noun wording ("voter") or a 10-item questionnaire using the verb wording that refers to the act of voting as a behavior ("voting/to vote"). The scripts are nearly identical to those used in Experiment 3 by Bryan et al. (2011), with the following minor differences that do not alter the substantive meaning of the questions or the substantive interpretation of either treatment. First, subjects who participate in the survey experiment either 2 or 3 days prior to Election Day see questions that refer to the upcoming election as "Tuesday's election" (for Kentucky, Mississippi, and Houston) or as "Saturday's election" (for Louisiana). Second, the wording of the response options provided for the last two questions in the 10-item questionnaire are altered to be more specific than in both Bryan et al. (2011) and Gerber et al. (2016). The full text of the treatment scripts may be found in Online Supplemental Appendix C.

Randomization procedure. Subjects are assigned to the noun condition or to the verb condition with equal probability using a simple random assignment procedure that each vendor implemented separately. Across both samples, 1579 (51.3%) were assigned to the noun condition and 1499 (48.7%) were assigned to the verb condition. Because the randomization procedure occurred by vendor, we conduct a randomization check separately for each sample by regressing treatment assignment on a vector of pre-treatment covariates and test the null hypothesis that these covariates are not jointly prognostic of treatment assignment. We fail to reject the null hypothesis for each sample and infer that the randomization procedures are valid.[14]

Outcome measurement. As in Bryan et al. (2011), our outcome variable was actual turnout

[14] Appendix Table A1 presents a regression table summarizing the randomization check and Appendix Tables A2 and A3 present unweighted and weighted balance tables, respectively. We observe imbalances on selected covariates between the noun and verb conditions. In the YouGov subsample, the noun group is more male, more Black, and less likely to have voted in the 2012 general election than the verb group. In the SSI subsample, the noun group is more likely to have voted in the 2004 general election and in the 2015 primary election and less likely to have voted in the 2006 general election.

in the 2015 general election, a behavior, measured using voter files. For subjects from the SSI sample, we do this by matching personal identifying information (name, year of birth, and zip code) to state voter files. All subjects in the SSI sample were successfully matched to state voter files after the experiment by a private vendor. For subjects from the YouGov sample, turnout data are supplied by the vendor. For all subjects, the turnout variable is coded 1 if the subject voted in the 2015 general election and 0 if the subject did not vote.

Estimation. We conduct our main analysis using the pooled sample and estimate the average effect on turnout of assignment to the noun condition on turnout, relative to assignment to the verb condition.[15] Our primary specification is an ordinary least squares model regressing a binary indicator for voting in the 2015 election on assignment to the noun block and pre-treatment covariates. Covariates include subject-level covariates such as the subject's race, gender, and past voting history; the number of days between the day the treatment is assigned and Election Day; state, date, and vendor fixed effects; and state-by-covariate interactions.[16]

Finally, we assess the robustness of our main analysis in several ways. First, we estimate noun wording effects among the subset of subjects who entered the experiment either on or the day before Election Day. Second, we estimate noun wording effects only among subjects in Kentucky and Louisiana, which are the two contexts where the election is both high salience and competitive ex ante, and therefore most likely to provide a context in which the psychological process of interest can occur. Third, we estimate noun wording effects among subjects in Kentucky and Louisiana

[15]While prior studies have also examined the effect of noun wording on stated interest in voting (operationalized as subjects' stated response to the survey items in each condition), we focus primarily on the effect of noun wording on the behavioral outcome, turnout, in this paper. We justify specifying the intent-to-treat effect as a substantively meaningful quantity of interest because nearly all subjects took up the treatment condition to which they were assigned and because failure to complete the entire survey is plausibly post-treatment and affected by treatment assignment. Among subjects assigned to the noun wording condition, 1557 (98.6% of 1579) viewed and responded to all 10 questions in the noun-wording block. Similarly among subjects assigned to the verb wording condition, 1476 (98.5% of 1499) viewed and responded to all 10 questions in the verb-wording block. We also observe virtually identical levels of treatment compliance in each subsample by survey vendor. Alternatively, one could estimate the Complier Average Causal Effect (CACE) by scaling the ITT by the estimated proportion of Compliers, where compliance is defined as viewing and responding to all 10 questions in the treatment arm to which one is assigned and viewing and responding to none of the questions in a treatment arm if not assigned tot that arm. Under this framework, we estimate the proportion of Compliers is about 98% and the ITT closely approximates the CACE.

[16]We additionally estimate treatment effects using a differences in proportions estimator with no covariates. See Table 3 for results. Results are not affected by using alternative estimators.

who entered the experiment on and the day before Election Day. Fourth, we assess whether the main results are sensitive to restricting the sample to those who viewed and completed all 10 survey items from their assigned treatment arm. While doing so may introduce post-treatment bias if treatment receipt is affected by treatment assignment, we nonetheless present this analysis to allay potential reviewer concerns about the sensitivity of our main result.

3 RESULTS

Does assignment to a 10-item online survey questionnaire that frames a respondent's potential voting behavior using noun wording ("voter") instead verb wording ("voting/to vote") affect turnout levels? As Table 3 shows, the turnout levels in the noun and verb wording groups are statistically indistinguishable. In the pooled sample, turnout levels are about 0.7 percentage points higher in the noun condition than in the verb condition (44.5% vs. 43.6%; regression-based treatment effect estimate=0.007; s.e.=0.012; n=3078; p=0.57, two-tailed).[17] This result is qualitatively similar when we partition the data by vendor or by state; the difference in turnout levels between the noun and verb conditions is both small and not statistically distinguishable from zero across all subsamples.

[Table 3 about here]

We now assess the robustness of our results by imposing a series of additional tests. First, we address a potential objection to our findings that our experimental design implements the treatment in the four days leading up to and including Election Day whereas the design in Bryan et al. (2011) implements the treatment in the two days leading up to and including Election Day. The effect reported by Bryan et al. (2011) may therefore exist only among subjects who receive their assigned treatment either on or the day before Election Day, but estimating treatment effects among all subjects in our study could mask this effect by calculating a weighted average of the noun effect among those treated just before the election and the noun effect among those treated earlier. To address this concern, we re-analyze the experiment only among subjects who entered the experiment and

[17] We refer the reader to Appendix Table A4 in the online Supplemental Appendix for full estimation results from the main analyses.

were assigned to either the noun or verb survey block either on or the day before Election Day. As shown in Appendix Table A5, our main result that noun wording has no effect on turnout is robust to this sample restriction (estimate=0.008; s.e.=0.018; n=1566; p=0.67, two-tailed).

Second, we address another potential objection to our study that our replication experiment includes electoral contexts without competitive elections (Mississippi) and without high-salience electoral contests (Houston). Consequently our main estimate may be underestimating the effect of noun wording on turnout in high salience, competitive electoral contexts where the social identity of being a voter may plausibly matter to affect political behavior. To address this concern, we partition the sample and estimate noun wording effects only among subjects in Kentucky and Louisiana. Appendix Table A6 reports that our main result that noun wording has no effect on turnout is robust to this sample restriction (estimate=0.0007; s.e.=0.014; n=2074; p=0.96, two-tailed).

Third, we address both of the above potential objections and re-analyze the experiment restricting the sample to subjects in Kentucky and Louisiana who enter the experiment on or the day before Election Day. Doing so provides a strong test of whether the hypothesis that noun wording affects turnout, because we are only examining the effect in places where high-salience and competitive electoral conditions exist, which Bryan et al. (2016) argue are the conditions needed for the psychological phenomenon of interest to occur. Appendix Table A7 shows that our main result that noun wording has no effect on turnout is also robust to both of these sample restrictions (estimate=-0.013; s.e.=0.021; n=983; p=0.54, two-tailed).

Finally, we show in Appendix Table A8 that the null result holds even when conditioning on subjects who view and respond to all 10 survey questions to which they are assigned. Among subjects who fully take up their assigned treatment condition, they noun wording has no effect on turnout (estimate=0.006; s.e.=0.012; n=3033; p=0.62, two-tailed). As this is not a random partition of the sample, this quantity is observational and presented only to allay potential concerns about whether our main result holds among this particular subgroup.

4 DISCUSSION

This article contributes to an ongoing debate in political psychology about whether small wording differences have outsized effects on behavior, and specifically to the debate whether mobilization appeals that include subtle linguistic cues linking desired behavior (political participation) to a person's social identity using a predicate noun (e.g., a "voter") instead of a verb (e.g., "one who votes") are able to induce dramatic increases in political participation rates by priming that identity.

Following the original experiment by Bryan et al. (2011) that reported large noun wording effects on turnout levels (by about 11 to 14 percentage points), a subsequent study by Gerber et al. (2016) instead found no effect, which casts some doubt on the robustness of the original finding. A rejoinder by Bryan et al. (2016) provided some possible explanations for the contrasting findings. They noted that some of the procedures in Gerber et al. (2016) differed from those in Bryan et al. (2011) and encouraged subsequent replication experiments to be conducted in competitive and important electoral contexts that allow subjects to plausibly experience the psychological phenomenon they argue exists. The experiment we conducted in 2015 is designed to directly examine some of these explanations for the null result in Gerber et al. (2016) by delivering identical noun and verb treatments using an Internet-based survey (as was done by Bryan et al. (2011)) and in competitive general election contexts.

We find no evidence that noun wording increases subjects' turnout. This null result is consistent across states and across survey vendors that use different procedures to recruit subjects and to match respondents to the voter file. Even when we partition the data to focus on subjects in competitive and high-salience contexts and subjects who receive the intervention either on or the day before the election, we find that the noun wording condition has no effect on turnout when compared to the verb wording condition. Furthermore, the sample size of this replication experiment is large and adequately powered to detect small effects,[18] which suggests that the null result is not likely due to sampling variability.

[18] We contrast the size of our experiment (n=3078) against those from the experiments from Bryan et al. (2011). Our sample is about 35 times as large as the Bryan et al. (2011) sample in California (n=88) and about 14.5 times as large as the Bryan et al. (2011) sample in New Jersey (n=214).

Our findings provides substantial evidence that a subtle linguistic prime that uses a predicate noun to prime people to think of themselves as voters has no larger effect on political participation than a verb prime, even in high salience and competitive electoral contexts. Figure 1 presents our main estimates alongside comparable estimates of the effect of noun wording (as compared to verb wording) on turnout from Gerber et al. (2016) and Bryan et al. (2011). Taken together, this growing body of evidence suggests that a subtle linguistic prime referring to a citizen as a voter is not generally sufficient to activate their social identity as a voter and increase the probability they vote.

[Figure 1 about here]

There are some important limitations to our findings. First, there may be subtle differences in the context, treatment, subject pool, or other features of the Bryan et al. (2011) study that account for the difference in results. Second, our finding does not reject the psychological theory that priming a person's social identity as voter can increase turnout. Further, and importantly, it may be that only a very limited subset of citizens hold a strong social identity as a voter that can be effectively primed to induce political participation. Alternatively, it may also be the case that a stronger intervention (rather than a subtle wording change in an online survey) is needed to prime one's social identity as a voter in a way that leads to behavioral change. We argue that future work assessing the effects of interventions that aim to increase turnout by priming a person's social identity as a voter could make progress by assessing heterogeneous effects by the strength of subjects' baseline social identity as a voter, by designing and testing stronger interventions to prime subjects' identities as voters, and by conducting manipulation checks to verify that subjects' construe the treatment as intended by the research and to verify that the social identity the intervention is attempting to prime is successfully primed. More generally, this study points to the need for careful and amply powered replication experiments to adequately assess the reproducibility of empirical claims that minor linguistic priming interventions affect political behavior, which are commonplace in the political psychology and "wise intervention" literatures.

REFERENCES

Bryan, C. J., Walton, G. M., and Dweck, C. S. (2016). Psychologically authentic versus inauthentic replication attempts. Proceedings of the National Academy of Sciences of the United States of America, 113(43):E6548.

Bryan, C. J., Walton, G. M., Rogers, T., and Dweck, C. S. (2011). Motivating voter turnout by invoking the self. Proceedings of the National Academy of Sciences of the United States of America, 108(31):12653–12656.

Gelman, S. A. and Heyman, G. D. (1999). Carrot-eaters and creature-believers: The effects of lexicalization on children's inferences about social categories. Psychological Science, 10(6):489–493.

Gerber, A. S., Huber, G. A., Biggers, D. R., and Hendry, D. J. (2016). A field experiment shows that subtle linguistic cues might not affect voter behavior. Proceedings of the National Academy of Sciences of the United States of America, 113(26):7112–7117.

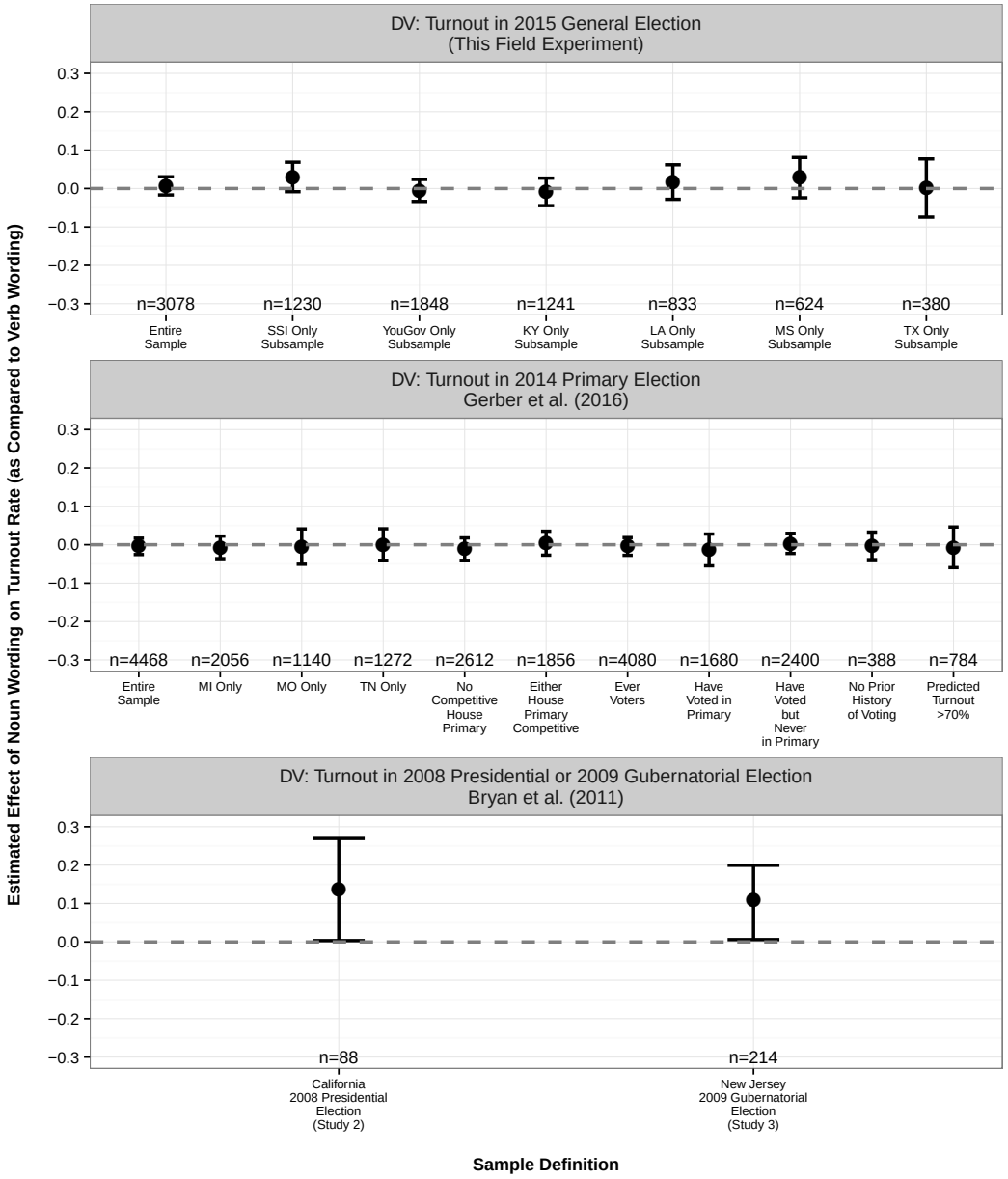
Green, D. P. and Gerber, A. S. (2015). Get Out the Vote: How to Increase Voter Turnout. Brookings Institution Press, Washington, D.C., 3rd edition.

Walton, G. M. (2014). The new science of wise psychological interventions. Current Directions in Psychological Science, 23(1):73–82.

Walton, G. M. and Banaji, M. R. (2004). Being what you say: The effect of essentialist linguistic labels on preferences. Social Cognition, 22(2):193–213.

FIGURE

Figure 1: Estimates of the Effect of Noun Wording on Turnout Rates as Compared to Verb Wording Across Studies. The figure presents estimates of noun wording treatment effects on turnout rates (as compared to the verb wording condition) with 95% confidence intervals and sample sizes. The top panel presents estimates from the covariate adjusted OLS model with inverse probability weights for this study (a full regression table is presented in Appendix Table A4 online). The middle panel presents covariate adjusted estimates reported from Table 1 in Gerber et al. (2016). The bottom panel presents non-parametric effect estimates from Studies 2 and 3 of Bryan et al. (2011). Details of how the 95% confidence intervals are calculated for the estimates from Bryan et al. (2011) are described in online Supplemental Appendix E.



TABLES

Table 1: Winning vote margins across empirical settings in Bryan et al. (2011), Gerber et al. (2016), and this field experiment. Vote margins are in percentage points.

Study	Subsample and Election Context	Winning Vote Margin (% pts.)
Bryan et al. (2011)	California 2008 presidential election	24.1
	New Jersey 2009 gubernatorial election	3.6
Gerber et al. (2016)	2014 midterm Congressional primary elections in Michigan, Missouri, and Tennessee	
	- Mean (of district-primary-level margins) among all district-primaries	62.6
	- Mean among contested district-primaries	33.0
	In districts where either Congressional primary is coded as competitive ex ante	
	- Mean among all district-primaries flagged as competitive	31.0
	- Mean among contested district-primaries flagged as competitive	24.7
	- Mean among all district-primaries	53.7
	- Mean among contested district-primaries	22.9
	In districts where neither primary is coded as competitive ex ante	
	- Mean among all district-primaries	67.0
	- Mean among contested district-primaries	38.4
	Michigan	
	- Mean among all district-primaries	65.8
	- Mean among contested district-primaries	26.2
	Missouri	
	- Mean among all district-primaries	61.8
	- Mean among contested district-primaries	38.9
	Tennessee	
	- Mean among all district-primaries	58.2
	- Mean among contested district-primaries	35.4
This field experiment	Kentucky 2015 gubernatorial general election	8.7
	Louisiana 2015 gubernatorial general runoff election	12.2
	Mississippi 2015 gubernatorial general election	34.1
	Houston TX 2015 mayoral general election	6.0

Table 2: Summary of prior studies.

Citation	Study	Treatment Arms	Treatment Delivery	Election	Geographic Restrictions	Screening Criteria	Total Study N	N in Noun and Verb Conditions	Outcome Variables	Mean Turnout Level among Verb Condition	Reported Effect of Noun Wording on Turnout
Bryan et al. (2011)	Experiment 1	Noun block or verb block	Internet survey	Nov 2008 election	California	Eligible to vote, not registered to vote, native English speakers, recruited via social networking website	34	34	Stated interest in registering to vote	N/A	N/A
	Experiment 2	Noun block or verb block	Internet survey	Nov 2008 election	California	Registered to vote, had not already voted (e.g., by mail) in election, native English speakers, recruited via university-administered online participant pool on social networking site	88	88	Turnout as measured using matched voter file records	0.818	0.137
	Experiment 3	Noun block or verb block	Internet survey	Nov 2009 election	New Jersey	Members of a randomly sampled and nationally representative panel administered by Knowledge Networks who were New Jersey residents and registered to vote in New Jersey at the time of the study	214	214	Turnout as measured using matched voter-file records	0.79	0.109
Gerber et al. (2015)	Replication experiment	Noun block, verb block, placebo, standard GOTV message	Telephone survey	2014 primary elections	Michigan, Missouri, and Tennessee	Registered voters with valid voter file records and who could be contacted by phone	11099	4468	Turnout as measured using matched voter-file records (all treatment arms)	0.311	-0.01

Table 3: Estimated Effect of Noun Wording (versus Verb Wording) on Turnout in the 2015 General Election. A full set of regression estimates may be found in Appendix Table A4 in the online Supplemental Appendix.

	Difference in Proportions (Noun Minus Verb)		Regression Estimate of Difference (Noun Minus Verb)		Verb Group	Sample Size	
Sample	Estimate	(SE)	Estimate	(SE)	Mean Turnout	Noun	Verb
Entire Sample	-0.018	(0.018)	0.007	(0.012)	0.438	1579	1499
YouGov Only Subsample	0.001	(0.023)	-0.005	(0.015)	0.518	960	888
SSI Only Subsample	-0.041	(0.027)	0.030	(0.020)	0.318	619	611
KY Only Subsample	0.006	(0.028)	-0.009	(0.018)	0.497	626	615
LA Only Subsample	-0.028	(0.033)	0.017	(0.023)	0.344	445	388
MS Only Subsample	-0.067	(0.040)	0.028	(0.027)	0.396	305	319
TX Only Subsample	-0.008	(0.051)	0.001	(0.038)	0.515	203	177