[Over?] Simplifying Direct Democracy

Effects of Additional Information on Ballot Measure Voting

W241 - Final Project - Spring 2020

Lorrel Plimier, Tim Slade, Padma Sridhar

Abstract

Citizens casting votes in support of or opposition to ballot measures is an exercise in direct democracy. However, the text provided to voters on the ballot can be difficult to understand. Can providing voters with additional information enable them to cast votes that better align with their preferences? This study, conducted online with a convenience sample of 441 adults, assessed the effect of presenting respondents with additional information about three ballot measures on their support for each measure. The additional information was operationalized as tables detailing the ballot's proponents and opponents, as well as the names and financial contributions of the top donors on either side of the issue. We did not find a statistically significant effect of the treatment. Not surprisingly, the degree of support was strongly predicted by the content of the measure and the respondent's party affiliation. The results suggest a study designed to have 80% power to detect an effect at $p < 0.05$ could require up to 450 respondents per treatment condition for two of the measures and over 409,000 respondents for the third.

Introduction

Direct democracy is the process that allows an electorate the power to vote legislation into law. An unstated assumption of direct democracy, however, is that the voters exercising this power fully comprehend the text of the legislation and are able to cast a vote that aligns with their preferences. Voters tend to vote in a manner that is consistent with their underlying values, beliefs and partisanship (Mendelsohn 1996). Because the United States is a representative democracy, voters here are accustomed to relying on their representatives to enact legislation on their behalf. They often rely heavily on policy platforms and party labels— indicators of values and beliefs—when voting for these representatives. In direct democracy, the information provided to voters on the ballot itself is stripped of any policy platform or party label and without these tools, voters are more likely to cast ballots that are not in line with their own preferences (Bowler & Donovan 1994; Burnett 2019). Providing additional information to voters, beyond what is presented on the ballot, is one way of simplifying the decision making process for them thus providing them the opportunity to vote in a manner that is aligned with their own values and beliefs.

There have been numerous observational studies, though few, if any, randomized experiments that explore the effects of voter education, ballot length, readability and position on the outcome of a ballot measure. These studies mainly rely on pre-election poll data, election results and post-election survey data. There is general agreement among the authors of the various studies that longer and harder to understand ballots lead to more drop-off (skipping questions on the ballot when voting) and more “no” votes (Bowler et al. 1992; Dubin & Kalsow 1997; Matsusaka 2012). The “no” vote is considered an acceptance of the status quo and is an indication that the ballot measure does not provide enough incentive to overcome this status quo (Bowler et al. 1992).

We hypothesized that one way to increase the incentive for voters to vote their preferences rather than defaulting to the status quo is to provide them with educational materials that lead to a better understanding of a ballot measure. Supplying the voter with a list of people and organizations that have endorsed or made a financial contribution to the campaign for the ballot measure can provide the voter with cues about whether the ballot measure is aligned with their values, beliefs and partisanship and thus whether they would prefer to vote for or against the measure. It is the effect of this provision of information to a voter that we study in this experiment.

Experimental Design

Our goal with this experiment was to explore the effect on voters' support for a ballot measure of supplementing the ballot's presentation with additional information. Manipulating an official election ballot would not only be unethical but also illegal, so we crafted our “ballot” in the form of a survey. Administering the survey over the internet permitted us to manipulate several features of the "voting" experience, which would allow us to explore the effect of additional information delivered in differing ways.

While we were interested in the effects of additional information on the support for any given ballot measure rather than on overall support across multiple ballot measures in a single election, we decided to include three measures on our ballot to make it more consistent with what a voter would experience in an official election. Every participant was presented with three summary ballot measures: one concerning parole laws (*parole*), one concerning ambulance company employees (*ambulance*), and one concerning kidney dialysis clinics (*clinic*).

Our initial design was extremely complex because we aimed to investigate the causal effect of multiple parameters of interest. (See ***Annex A*** for details.) Ultimately, we recognized it would be infeasible to administer 36 different treatment conditions across 18 blocking conditions within our existing time and resource constraints. We therefore developed the simplified design based on the treatment variant we theorized would have the greatest effect.

Our hypothesis was that providing lists of financial contributors and endorsers both for and against a ballot measure would change the level of support for that ballot measure. Half of the respondents were randomly assigned into the *control* condition and half into the *treatment* condition. Our survey presented the three ballot measures one at a time in either the treatment condition or the control condition and followed each presentation with a posttest measurement as depicted below in ***Table 1***. We observed three $Y_i(0)$ potential outcomes for each subject assigned to the control group and three $Y_i(1)$ potential outcomes for each subject assigned to the treatment group.

Table 1: *Posttest Control Group Experimental Design*

Treatment Group	R	X	O	X	O	X	O
Control Group	R	-	O	-	O	-	O

As illustrated in ***Figure 1***, respondents assigned to the control condition were presented with 1 the abbreviated summary of the ballot measure which appears on the actual ballot during an election, and 2 a slider they were directed to position so as to indicate their degree of opposition to ← or support for → the measure. The slider could be positioned anywhere along a 100-pt scale ranging from -50 (full opposition) to 50 (full support); however, respondents were given no visible indication of the numeric value associated with the position of their slider. All sliders were initialized at the 0 position (neutral), equidistant from either end of the scale.

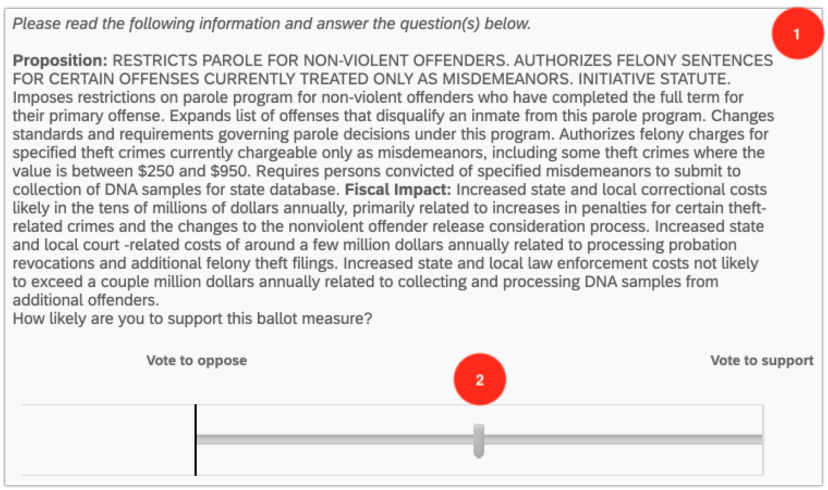

Figure 1: *Ballot measure as presented to respondents randomized into* Control *condition*

Respondents assigned to the treatment condition were presented a slightly different prompt, depicted in ***Figure 2***. The summary of the ballot measure 1 and the slider 2 presented to the control respondents were retained. However, treatment respondents were additionally shown a table of organizations who had publicly announced their endorsement in support of or opposition to the measure 3 as well as tables indicating the top financial contributors on either side and the magnitude of their contributions. 4

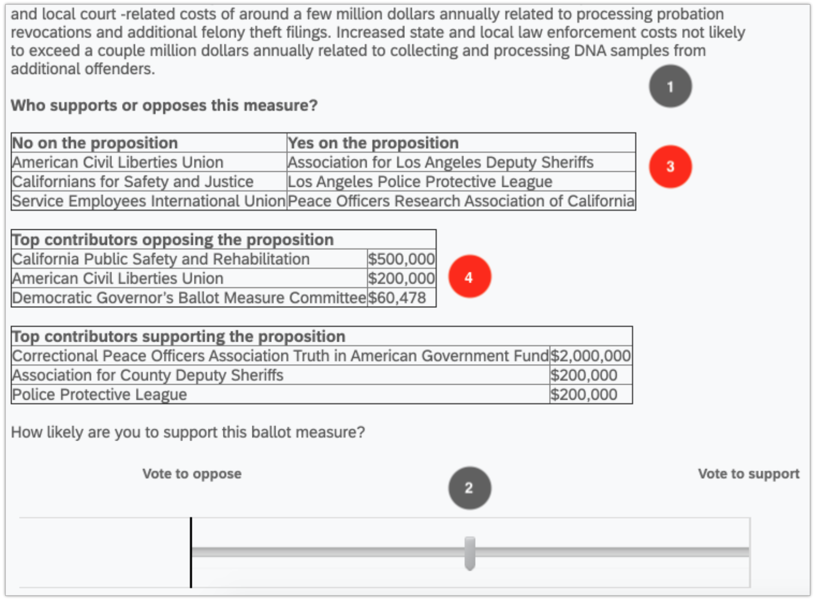

Figure 2: *Supplementary information presented to respondents randomized into* treatment *condition*

Figures 1 and ***2*** are based on the *parole* measure; the same format was retained for the *ambulance* and *clinic* measures. Respondents were presented with all three ballot measures, and their *control* or *treatment* assignment was persistent across all measures. We randomized the order of presentation for the ballot measures in order to eliminate any bias there might be in the presentation order. We did not treat the ballot order as a separate treatment condition, however, instead relying on Matsusaka (2012), a randomized natural experiment that found that the order of the ballot measure on a ballot had no effect on the outcome of the vote for that measure.

In presenting the survey, we wanted to block the participants by political party and education level. We theorized that one's political party would likely be correlated with the outcome of the vote on the ballot measures and that this would be especially true when presented with lists of partisan financial contributors and endorsers. As media coverage of politics often emphasizes that level of education is an important predictor of polling outcomes and voting behavior, we chose to block on that as well. For the purposes of blocking, we collapsed the political party into three categories: Republicans, Democrats, and everything else, which included both Independents and an "other" category that would give participants the opportunity to enter their own party description. We similarly collapsed the five levels of education into two categories: "less than college degree" and "college degree or more".

Design Considerations

For the purposes of our study, we used the services offered by Qualtrics to implement both the pilot and final design. Using Qualtrics had a few key benefits:

- The survey flow was configured to randomize subjects easily at a specified point in the survey. We chose this point to be upon launch of the survey for the participant.
- Blocking was configurable by question, which enabled us to quickly set up blocks by political affiliation and education.
- The Qualtrics randomization not only ensured that subjects would be equally divided into treatment and control but also across our specified blocks.
- Qualtrics allowed us to embed information in the survey to capture additional metadata such as the IP Address, whether or not they finished the survey, a unique Response ID for each participant, and the time they spent on each question.

We required that all subjects be at least 18 years of age and give consent to participating in the survey. Aside from optionally submitting an email address for a gift card raffle, participants were required to answer all of the questions on a page before proceeding to the next page, ensuring we would get a complete set of responses from those who finished the survey. We also prevented participants from returning to previously answered questions to update their responses.

Table 2: *Potential 1st-degree reach for venues in which the survey link was posted.*

Facebook	3,969
LinkedIn	1,361
Twitter	179
Social Media Subtotal	5,509
Slack	8,265
Reddit	129,000
Total	**142,774**

Qualtrics provides the option to embed information about the inbound source of the link which launches the survey. This allowed us to track which survey results were generated by our contacts on social media and which were generated by participants from outside our individual networks. We sourced participants from LinkedIn, Twitter, Facebook, direct communication with family and friends, several Slack workspaces including the UC Berkeley School of Information

Slack community, and the Reddit r/SampleSize forum. The potential first-degree reach for our survey in these venues is reported in ***Table 2***.[1]

Pilot Study

We perused the peer-reviewed literature for similar studies that could help us make an *a priori* estimate of the likely effect size. We were unable to find any such studies. We therefore conducted a very small pilot study with two goals: (1) to obtain feedback on the survey's presentation and format from actual respondents, and (2) to estimate a plausible treatment effect we could use to run power calculations. We recruited 31 family members and friends and randomly assigned them to treatment and control groups as described above. We did not block on any of our covariates for the pilot study.

Figure 3 shows the distribution of support for each of the three ballot measures in our pilot study. The treatment appears to pull respondents away from the center line at zero (neither supporting nor opposing the measure) toward the outer edges. There are clear spikes at the most extreme values of -50 (strong opposition) and 50 (strong support).

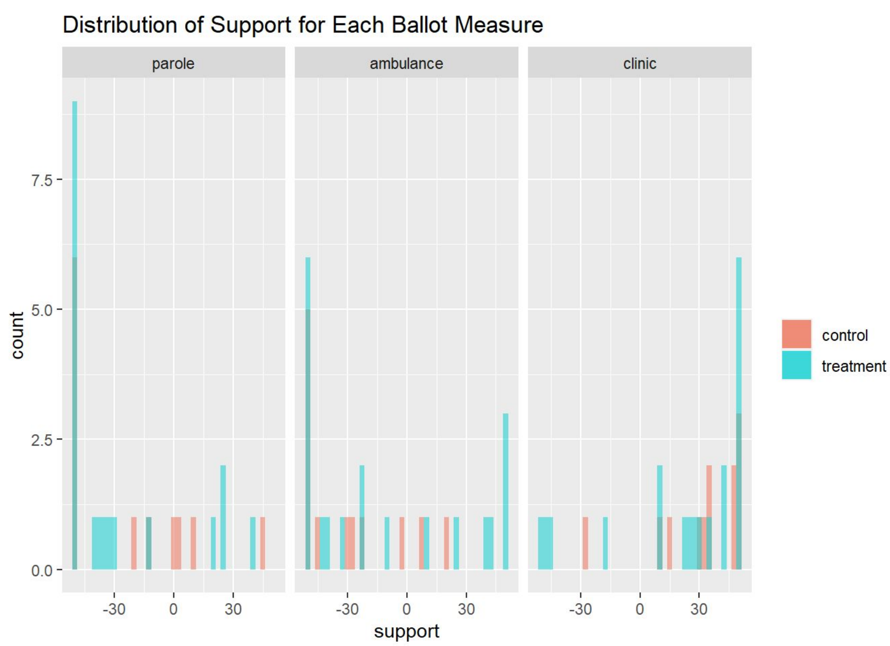

Figure 3. *Distribution of support for each ballot measure during the pilot study. The upper and lower bounds lead to left skewed or right skewed distributions depending on the content of the ballot measure.*

[1] If all first-degree contacts and no second-degree contacts in each of those venues had been exposed to our survey recruitment messages—which we acknowledge is highly improbable due to network effects, social media usage patterns, and algorithmic presentation of content—our 441 respondents would represent a response rate of approximately 4.6% (256/5509) for social media and 0.13% (184/137,265) for Slack and Reddit.

In the pilot study, both the *ambulance* and *clinic* measures had a substantially larger effect size than the *parole* measure. ***Table 3*** shows the effect size for each of the three measures and an estimate of how many respondents we would need in each treatment condition to reach 80% power at that effect size. We used these estimates to set a goal of 250 participants for each of our treatment groups.

Table 3: *Effect size and sample size required to reach 80% power as estimated from mean support observed during pilot study (n = 31)*

Measure	Mean Support Treatment	Mean Support Control	Effect Size	N to reach 80% Power
parole	-27.68	-22.75	0.158	316
ambulance	-10.63	-29	0.505	33
clinic	20.47	31.16	0.345	68

Analysis

Survey Flow and Attrition

Figure 4 provides an overview of how the 441 respondents proceeded through our survey. Participants were presented with 12 questions divided into eight sections as follows: *Consent* (1), *Demographics* (7), *Disclaimer* (1), *Ballot 1* (1), *Ballot 2* (1), *Ballot 3* (1) and *Email* (1). Only the email prompt was optional. Randomization into treatment or control was assigned at the point of loading the survey webpage, but it was only after the *Disclaimer* that the treatment and control participants received different questions. We refer to submission of the *Disclaimer* question as the *point of divergence*.

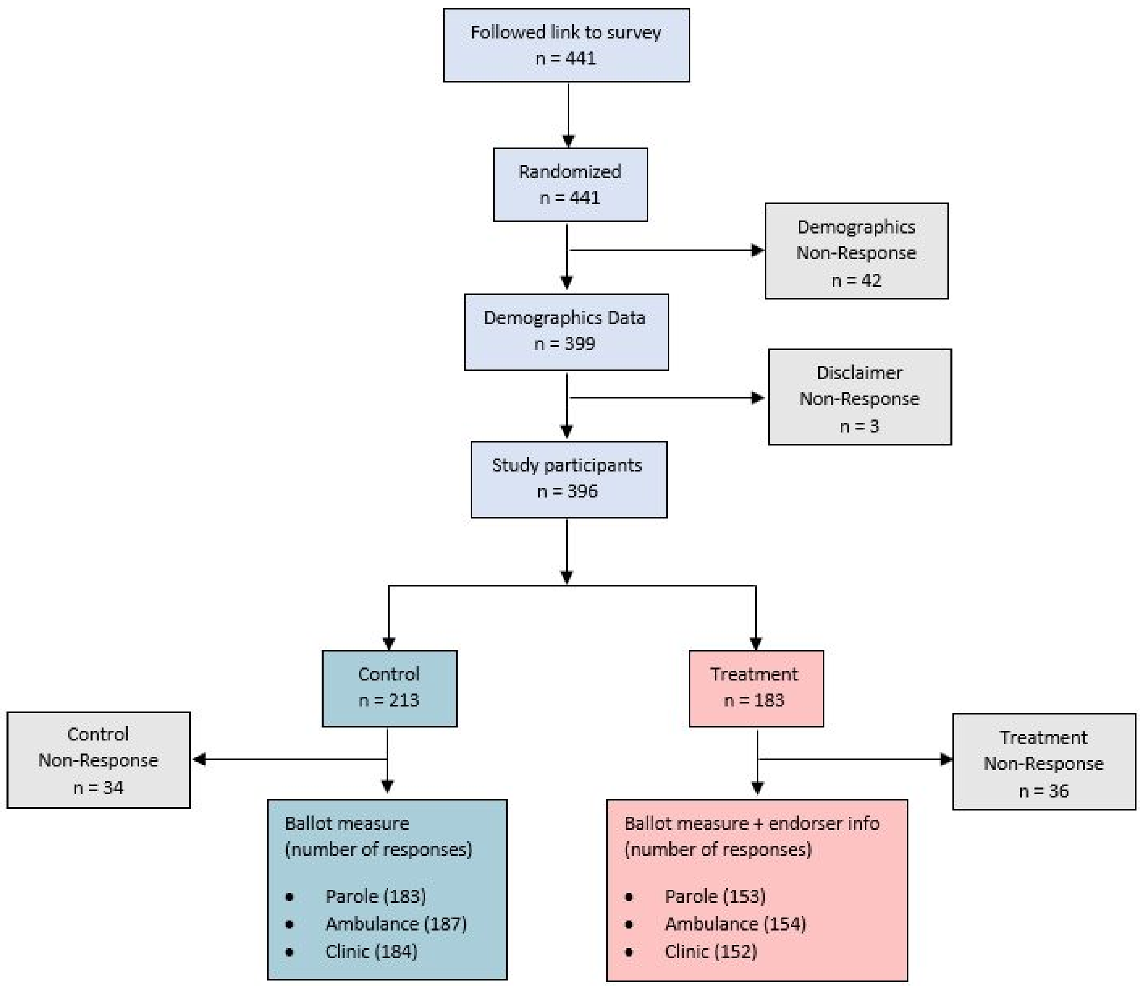

Figure 4: *Consort diagram depicting the flow of respondents through our survey. Overall attrition was 26% (115 of 441), evenly split between treatment and control (p = 0.052).*

The majority of our attrition took place prior to completion of the initial set of seven *Demographics* questions. A few more dropped out without completing the *Disclaimer* acknowledgement, with the remainder attriting from the *Control* condition (34) or the *Treatment* condition (36) over the course of the three ballot measures. We limited our differential attrition analyses to those attritors who dropped out after the point of divergence, as those who dropped out before shared a common experience; their attrition could not have been caused by the treatment assignment.

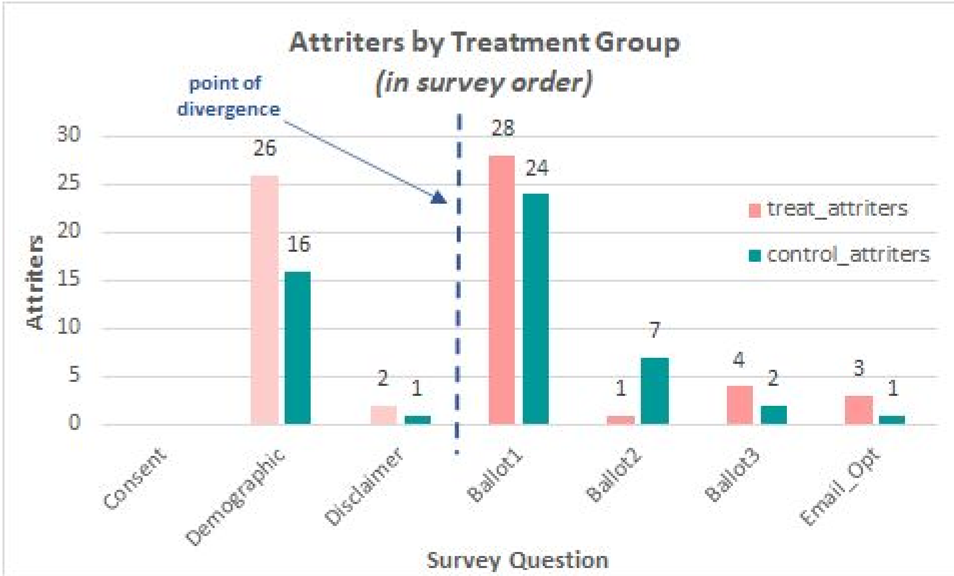

***Figure 5**: A substantial proportion of attriters abandoned the survey before their assignment to treatment or control had resulted in a divergent experience. Attrition was not found to be significantly different between treatment and control past the point of divergence.*

Figure 5 shows in detail the number of attriters between the treatment and control groups at each section in the survey, starting from the Consent and ending at the optional Email question. All 441 participants consented to the survey. We saw that 45 attritors did not complete the survey past the point of divergence. Of those who entered into their respective treatment and control conditions, 70 subsequently attrited at some point before the end of the survey. The rate of attrition for these 70 participants was generally consistent during the first week we launched the survey and declined gradually over the course of the 4 week period (See ***Annex B** - **Figure B1*** for details).

We had a total of 115 attriters across the board, resulting in an **overall attrition rate** of 26%. At the conclusion of the survey we confirmed with Qualtrics that there were no major outages that could have impacted the attrition numbers. Looking more closely at attrition by ballot, we see that 52 attrited at *Ballot 1*, the first ballot they received, regardless of whether that ballot was *parole*, *clinic*, or *ambulance*. A similar tally was made for *Ballot 2* and *Ballot 3* which had 8 and 6 attriters, respectively.

We performed differential attrition checks on the specific ballot measures, three possible ballot positions, and six experimental blocks (see ***Annex B** - **Tables B1-B3***). The difference in proportion of attriters between treatment and control was not statistically significant for 11/12 tests. The only significant result was for *Ballot 2* with a p-value of 0.04. However, the total number of attriters for *Ballot 2* across both treatment conditions was $n = 8$: too small to have any practical significance in the larger context of differential attrition.

Analysis of Demographics

Demographic information was collected for all survey participants. After consenting to participate in the survey, every participant was asked a series of seven questions regarding age, gender, education, income, political affiliation, and voting habits, covariates within which we thought heterogeneous treatment effects might be possible.

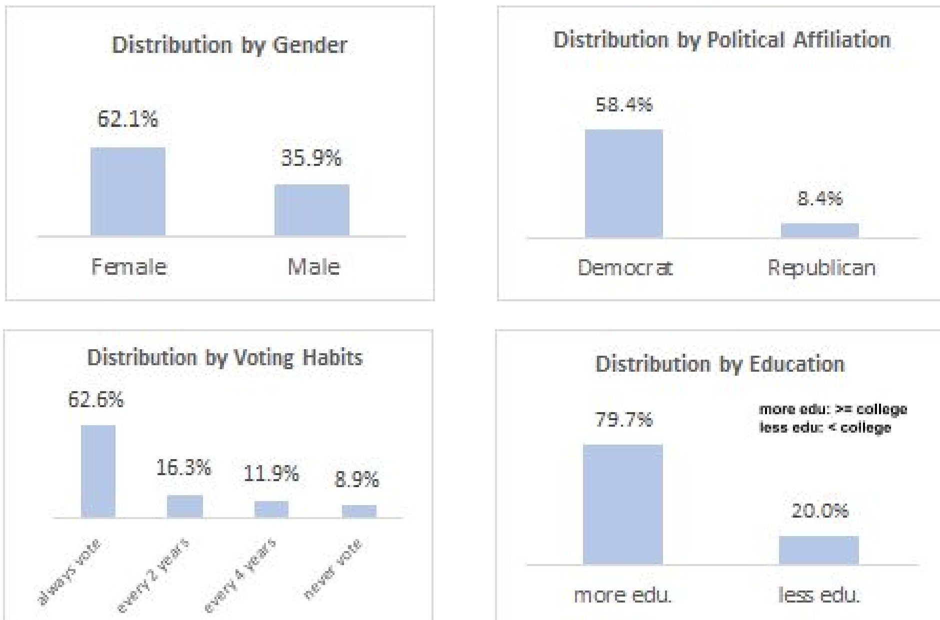

Figure 6: *Distribution of key demographic variables. Our respondent pool skewed female, overwhelmingly Democratic, highly politically engaged, and highly educated. The partisan imbalance proved to be a limitation in some later analyses.*

Figure 6 highlights the key differences in demographics for a population of *n = 404*, the number of survey respondents who fully completed this section of the survey. There are about 73% more female than male respondents. We also had a few who identified as Other/Nonbinary/Transgender but the *n* for this group was very small (< 6). We see an even greater imbalance between political affiliations, with only 8.4% of respondents identifying as Republicans while 58.4% identified as Democrats. The distribution by voting habits leaned in favor of those who “always vote”, indicating they participated in presidential, midterm, and local elections. And finally, over 79% of our respondents had *at least* a college degree. From a geographical standpoint, 93% of responses were from the United States and 32% from California alone. Over 60% of responses were leveraged via our personal social networks.

We used Welch’s two-sample t-test to explore the differences in means by demographic covariate for each of the ballot measures. This was done for the population *n = 396*, the number of survey respondents who fully completed the Disclaimer section of the survey and entered their respective control/treatment conditions (see ***Annex B*** ***- Table B4*** for details). The results were significant for support for the *parole* measure between more/less educated voters (*p = 0.0076*), support for the *clinic* measure between high/low income voters (*p = 0.012*), and support for the *parole* and *clinic* measures between Republicans and Democrats (*p = 0.000099 and 0.012, respectively*).

A covariate balance check was conducted for the variables of interest that were captured in the demographics questions to ensure there was an equal distribution between treatment and control groups. This was done for the *n=396* participants who crossed the point of divergence. The regression automatically excluded any missing values (participants that attrited later in the survey) resulting in the 377 degrees of freedom we see below.

```
Coefficients:
                                   Estimate Std. Error t value Pr(>|t|)
(Intercept)                       -0.123139   0.509332  -0.242   0.8091
age                               -0.000957   0.002295  -0.417   0.6769
genderMale                         0.092597   0.055840   1.658   0.0981 .
genderNonbinary/Transgender/Other -0.008259   0.232344  -0.036   0.9717
genderPrefer not to say           -0.504108   0.359773  -1.401   0.1620
education2: high school            0.681621   0.526520   1.295   0.1963
education3: some college           0.660449   0.514366   1.284   0.1999
education4: college degree         0.695034   0.515599   1.348   0.1785
education5: some post-grad         0.666440   0.517249   1.288   0.1984
education6: graduate degree        0.657867   0.516466   1.274   0.2035
income2: $50k to $100k            -0.120248   0.078217  -1.537   0.1250
income3: over $100k               -0.088775   0.073889  -1.201   0.2303
voting2: every 4 years            -0.051021   0.117085  -0.436   0.6633
voting3: every 2 years            -0.119817   0.109783  -1.091   0.2758
voting4: always vote               0.025777   0.100791   0.256   0.7983
partyIndependent                   0.051596   0.060750   0.849   0.3962
partyOther                        -0.093537   0.105941  -0.883   0.3778
partyRepublican                    0.091768   0.094176   0.974   0.3305
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4998 on 377 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.04036,    Adjusted R-squared:  -0.002916
F-statistic: 0.9326 on 17 and 377 DF,  p-value: 0.5354
```

Figure 7*: OLS regression results used to confirm covariate balance where the* `treatment` *variable is regressed on the covariates.*

As seen in ***Figure 7***, the difference between the treatment and control groups across all covariates and their distinct levels was not significant, suggesting our randomization and blocking procedures worked as expected. The analysis of variance resulted in a p-value of 0.53 indicating that this model was not statistically different from the model without the covariate terms and thus that these covariate terms failed to predict whether the participant was in the treatment or control group.

Modeling and Results

Our core interest was in determining whether receiving the treatment was predictive of a respondent's level of support for a ballot measure. Our approach, summarized in ***Table 4***, was to proceed systematically through four steps of modeling for each of our ballot measures of interest.

Our simple *base model* (1) regressed the degree of `support` for a measure on the respondent's `treatment` status. In keeping with the literature (Bowler, et al. 1992; Bowler & Donovan 1994; Burnett 2019) on the influence of key covariates such as `party` affiliation, `intensity` (of self-identification with a party), level of `education`, `income`, and `voting` habits on observable political behavior, however, we expected much of the support for each ballot measure to be explainable by those features whether singly or in combination. We therefore developed *covariate models* (2) that included each of those covariates as an additional regressor alongside the treatment. Doing so controlled for the influence of those covariates over the outcome; in the covariate models the coefficient on the treatment indicator thus represented the treatment effect that is independent of the influence of those other covariates. Next, we generated *interaction models* (3) by modifying each covariate model to include an appropriate interaction term allowing us to assess whether members of one of the groups defined by the covariates were more susceptible to the treatment effect. Finally, we developed a *long* model which included all of our key covariates and their interaction terms.

Table 4: *Modeling strategy*

(1)	*Simple model*	$S = \beta_0 + \beta_1 T$	β_1 : the influence of receiving the treatment on the outcome variable
(2)	*Covariate model for increased precision*	$S = \beta_0 + \beta_1 T + \beta_2 C$	β_2 : the influence of the covariate on the outcome variable independent of the treatment
(3)	*Interaction and conditional average treatment effects*	$S = \beta_0 + \beta_1 T + \beta_2 C + \beta_3 T * C$	β_3 : the heterogeneous treatment effects within groups defined by a covariate C
(4)	*Long model*	$S = \beta_0 + \beta_1 T + \beta_m X_i + \cdots + \beta_n X_j$	β_m to β_n : the heterogeneous treatment effects of covariates i through j $X_i \cdots X_j$: the covariates both as simple regressors and as interacted with treatment term T.

S = degree of support for the ballot measure: variously *parole*, *clinic,* or *ambulance*
T = Treatment (1) or Control (0) status
C = The covariate of interest: variously *party*, *intensity*, *education*, *income*, or *voting* habits
X = The covariate as regressor with accompanying interaction term

We illustrate this general approach using respondents' `support` for the *parole* measure as the outcome variable of interest and `party` as the covariate. ***Figure 8*** presents the regression results in a tabular fashion.

Predicting support for the parole measure

	Dependent variable:		
	Parole		
	Simple	Party as add'l regressor	Party interacted
	(1)	(2)	(3)
Treatment	-0.122	-1.141	-5.083
	(-6.093, 5.850)	(-6.864, 4.582)	(-12.484, 2.317)
Republican (Control)		32.191^{***}	26.184^{***}
		(21.353, 43.029)	(10.966, 41.401)
Independent (Control)		6.260	1.991
		(-0.403, 12.923)	(-7.392, 11.375)
Other Party (Control)		-4.089	-8.009
		(-16.021, 7.843)	(-22.731, 6.713)
Republican (Treatment)			12.545
			(-9.147, 34.237)
Independent (Treatment)			8.844
			(-4.492, 22.181)
Other Party (Treatment)			10.583
			(-14.593, 35.760)
Constant	-28.787^{***}	-32.198^{***}	-30.491^{***}
	(-32.816, -24.757)	(-36.622, -27.773)	(-35.360, -25.622)
Observations	336	336	336
R^2	0.00000	0.099	0.107
Adjusted R^2	-0.003	0.088	0.088
Note:			$^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

Predicting support for the parole measure

Base model is (a) Democrats in (b) control

Figure 8: *Support for the* parole *ballot measure predicted by* (1) *a simple linear regression model,* (2) *a model which includes* `party` *affiliation as a regressor, and* (3) *a model which includes* `party` *affiliation as both a regressor and interacted with* `treatment`*. The base condition is control* (`treatment = 0`) *and* `party = Democrat`.

The estimated coefficient for the `treatment` term varied significantly across models. Adding `party` as a regressor accounts for a meaningful amount of the variance in the score predictions (adjusted R^2 = 0.088), but only the `party = Republican` coefficient is statistically significant.

(It is also substantively meaningful, shifting support by approximately 26 to 32 points on a 100-point scale.) While including an interaction term between `party` and `treatment` alters the coefficients, it failed to improve the model's explanatory value (the adjusted R^2 remains at 0.088).

Figure 9 presents the same analysis graphically. The width of 95% confidence intervals on the estimated coefficient for `treatment` remained relatively constant across models, indicating that our additional covariates and interaction terms did not substantially improve our precision as we had expected. In all cases, the confidence intervals on `treatment` include 0, meaning we cannot be certain our treatment had an effect on our outcome.

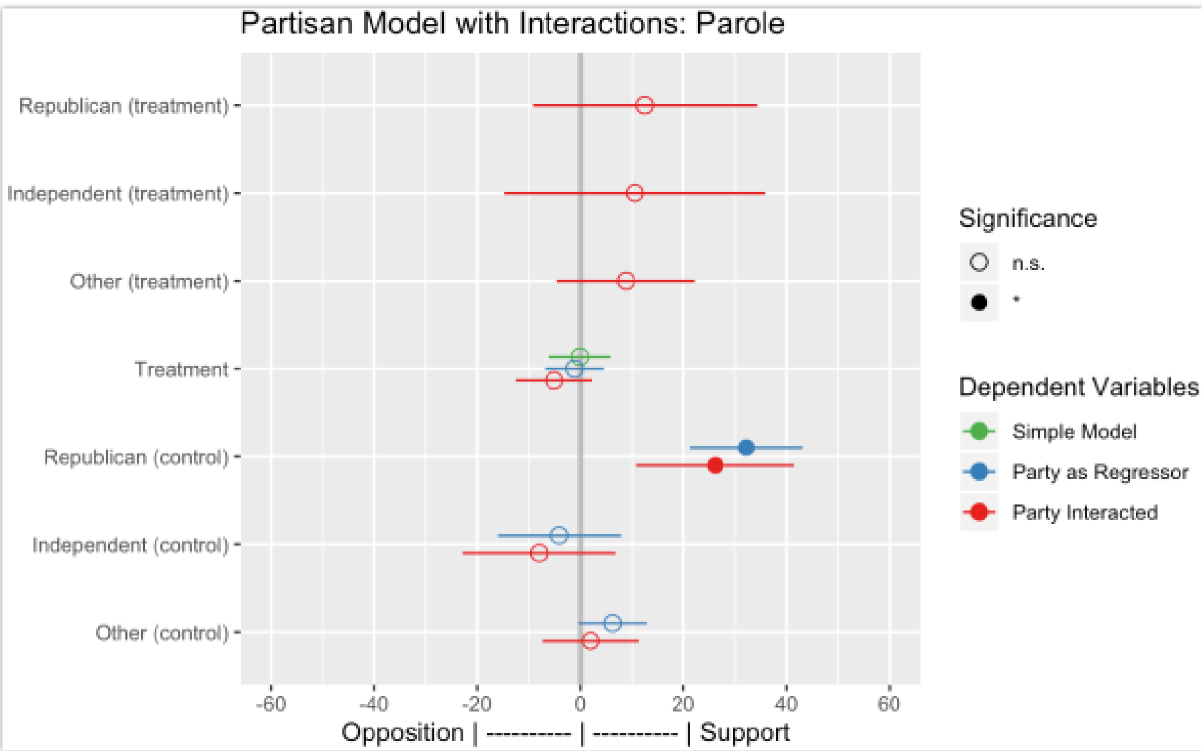

Figure 9: *Graphical presentation of the regression models presented in tabular fashion in Figure 8. The consistent width of the 95% confidence intervals for the* `treatment` *variable indicates the addition of* `party` *to the model as a regressor or in an interaction term with* `treatment` *did not substantially improve the precision of our estimates.*

We elide further discussion of these models for the sake of parsimony. Our key finding is simple: the additional information we provided regarding contributors and endorsers never had a statistically significant treatment effect across any of the models we developed and evaluated.

In a few cases, the coefficients on the covariates emerged as both substantively meaningful and statistically significant. Given the number of models and covariates tested, however, we cannot be certain these results were legitimate rather than the spurious result of multiple comparisons.

For instance, the policies and rhetoric of Republican leaders would suggest that `party = Republican` would be correlated with increased support for the *parole* ballot measure; as expected, the covariate model estimated the effect at $\beta = 32.19 \pm 10.8,\ p < 0.001$. Similarly, given a base population of Democrats, `intensity = Strong` could be theorized to correlate with increased opposition to the *parole* ballot measure, and the covariate model indeed estimated the effect at $\beta = -12.66 \pm 7.9,\ p < 0.01$. The influence of the `party` variable was generally expected—our experimental design blocked on party affiliation for precisely this reason—but what is to be made of `intensity = moderate` and `party = Republican` jointly predicting support for the *parole* $(\beta = -27.20 \pm 25.6,\ p < 0.05)$ and *clinic* $(\beta = -32.61 \pm 26.6,\ p < 0.05)$ measures, while no other combinations of partisanship and intensity were significant for any of the ballot measures? The most appropriate posture is to recognize that several of these analytical cells suffer from small *n* and that investigating effects on such a granular level requires more power than our experiment provides.

In search of additional power

A hypothesis test will more readily fail to reject the null hypothesis—even if the null hypothesis is incorrect—if the sample size is insufficiently large. We reasoned that if our treatment (providing information about financial contributors and endorsers) indeed had an effect, it should hold across ballot measures. If that were the case, we could reorganize our data to convert our ~396 observations split across three outcomes of interest (`support_parole`, `support_ambulance`, `support_clinic`) into 3 * 396 = 1188 observations on a single outcome of interest (`support`) as depicted in ***Figure 10***. We would need to use robust standard errors to adjust for the clustering of observations within respondents, but reasoned that doing so would incur a lower penalty than the gain we would obtain from the additional observations.

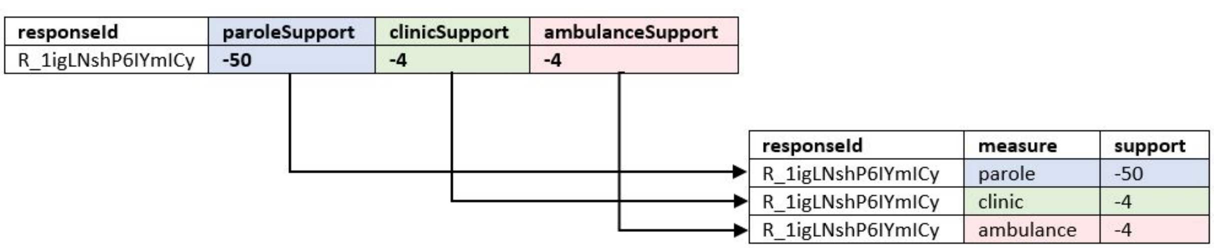

Figure 10: *Illustration of the process used to pivot the dataset from one observation per respondent with distinct support variables (left) to three observations with a single support variable (right).*

We applied the same modeling strategy as above, developing a *base model*, a model with *covariates*, and a model with *interaction* effects. ***Figure 11*** presents the results of this analysis.

Predicting support for any given measure

	Dependent variable:		
	Support		
	Simple	Measure added as regressor	Measure interacted w Treatment
	(1)	(2)	(3)
Treatment	-3.229	-3.132	-0.122
	(-7.712, 1.254)	(-6.852, 0.587)	(-6.115, 5.871)
Ballot Measure: Ambulance		23.329^{***}	25.782^{***}
		(18.721, 27.937)	(19.695, 31.868)
Ballot Measure: Clinic		50.026^{***}	51.662^{***}
		(45.784, 54.269)	(46.134, 57.190)
Treatment x Ambulance			-5.406
			(-14.705, 3.894)
Treatment x Clinic			-3.595
			(-12.182, 4.991)
Constant	-2.926	-27.416^{***}	-28.787^{***}
	(-5.899, 0.047)	(-30.807, -24.025)	(-32.740, -24.834)
Observations	1,013	1,013	1,013
R^2	0.002	0.320	0.321
Adjusted R^2	0.001	0.318	0.318
Note:			$^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

Predicting support for any given measure

The base conditions are treat='control' and measure='parole'

Figure 11: *Models predicting `support` using a pivoted dataset of three observations per respondent and robust standard errors due to clustering by respondent. Including which `measure` supplied the score adds significant explanatory value (adjusted R^2 = 0.318, greater than any other covariate we modeled) and reduces the effect of `treatment` to near zero.*

As the model confirms, respondents' degree of `support` was powerfully though not solely driven by the actual contents of the ballot measures: an already-small and statistically insignificant treatment effect is driven nearly to zero by the inclusion of a covariate representing the contents of the ballot measure on which they voted.

Once again, for the sake of parsimony, we omit extended discussion of the dozens of models that we tested on the basis of this expanded dataset. We determined that a significant coefficient for treatment can be extracted from a model that includes `party` and `intensity` as interacted covariates, but only under the naïve condition that excludes `measure`. As the exclusion of the contents of the ballot measure would create substantial omitted variable bias, we conclude this finding was spurious and do not report it.

Conclusion and Next Steps

Knowing we were underpowered in this experiment, we calculated the effect size of the treatment in each of our ballot measures and used this to estimate the sample size we would have needed to achieve 80% power with that effect size. ***Table 5*** shows the results for all three of our ballot measures.

Table 5: *Effect size and sample size required to reach 80% power as estimated from mean support observed during the full experiment (n = 396).*

Measure	Mean Support Treatment	Mean Support Control	Standard Deviation	Effect Size	N to reach 80% Power
parole	-28.91	-28.79	27.77	0.0044	409,261
ambulance	-8.53	-3.01	33.22	0.1664	286
clinic	19.16	22.88	28.33	0.1312	458

That the *parole* measure should require nearly one thousand times as many respondents as the *clinic* measure—at 409,261 respondents per treatment condition compared to merely 458—merits further review. As the effect size is one of the primary drivers of statistical power, we chose to investigate the issue further.

$$ES = \frac{\mu_1 - \mu_0}{\sigma}$$

As depicted in the equation above, the effect size is a function of two quantities: a difference in means in the numerator, and the standard deviation in the denominator. A small effect size could be driven by either a small difference in means or a large standard deviation. In the case of the *parole* measure, the difference in means was minuscule (a mere 0.12 points on a 100-point scale) while the standard deviation was substantial (27.77 points).

Figure 12 shows the distribution of the outcome variable for each of the three ballot measures. In both treatment and control groups, more people (68 in control and 53 in treatment) voted -50 to oppose the *parole* ballot measure than voted at any other position along the sliding scale. The remainder of the votes were spread out across the scale with a long tail in support of the ballot measure. In each of the treatment and control groups, 25 people voted in support of the measure somewhere along the scale between zero and 50 and nine of those voted at the far end of the scale at 50. This wide distribution generated a large standard deviation which when combined with the small difference in means resulted in such a small effect size.

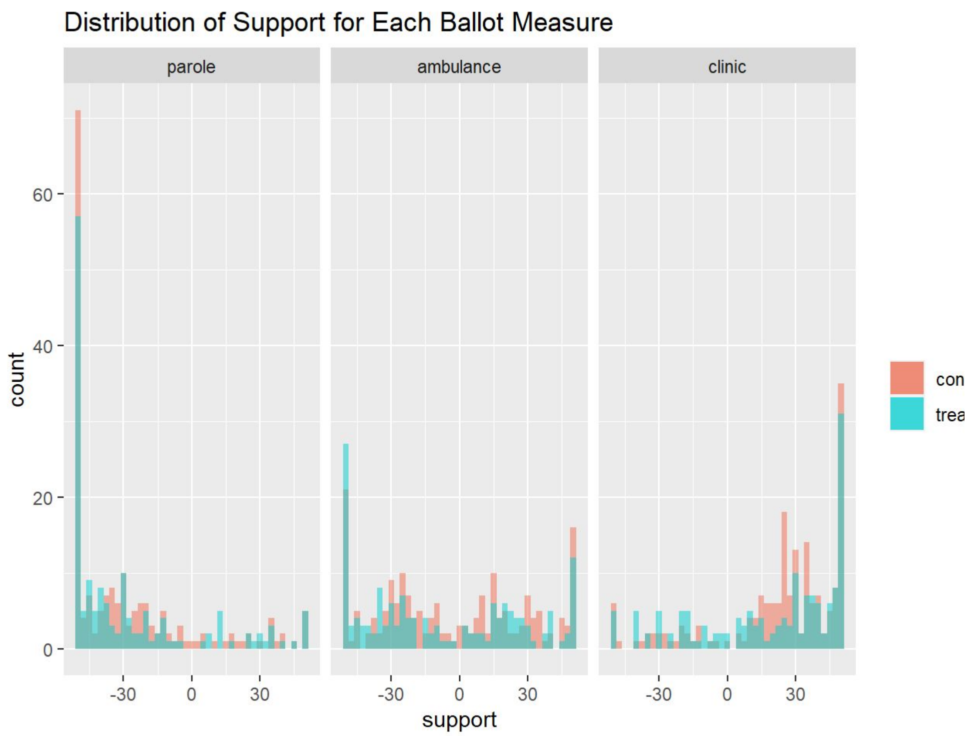

Figure 12. *Distribution of support for each of three ballot measures in the full experiment. The extreme left skew of the parole measure and the clustering of treatment and control led to a minute difference in means.*

As ***Figure 12*** depicts, the pattern of responses to the *parole* measure differed substantially from those of the other two measures. Having sought and failed to find analytical errors that could explain the differences, we reviewed the text of the measure and noticed several unique characteristics of the proposed legislation. First, the measure predicted a likely fiscal impact of increasing the state and local corrections costs by tens of millions of dollars annually; we theorize this would provide a strong incentive for fiscal conservatives to oppose it. Second, it would require DNA sampling for specified misdemeanors, something bound to alienate voters concerned about privacy interests. Finally, the measure would reduce parole eligibility and reclassify certain misdemeanors as felonies. Again, a large group of people—those opposed to the widespread harm caused by mass incarceration—would be inclined to vote against the *parole* measure. We theorize that these three key features of the *parole* ballot measure served to alienate a large proportion of the population.

We suspected that party affiliation would influence voters' support for ballot measures, and in our experiment it appears that it did. As our social networks skew highly educated, highly engaged, and Democratic, so too did our respondent pool. Unfortunately, temporal and financial constraints precluded a broader recruitment effort which could have attracted a more diverse set of respondents. Were we to use this experiment as a pilot for a larger experiment, we would focus on recruiting a sample that is more representative of the distribution of political affiliation in society. Overall, we hope this study provides continued motivation and encouragement for research into the impact of voter education on the practice of direct democracy.

References

1. Bowler, S., Donovan, T., and Happ, T., Ballot Propositions and Information Costs: Direct Democracy and the Fatigued Voter. *The Western Political Quarterly*, 559-68 (1992).
2. Bowler, S., Donovan, T. Information and opinion change on ballot propositions. *Polit Behav* 16 (1994) 411–435.
3. Bowler, S., Information availability and information use in ballot proposition contests: Are voters over-burdened? *Electoral Studies 38* (2015) 183-191.
4. Burnett, C.M., Information and direct democracy: What voters learn about ballot measures and how it affects their votes, *Electoral Studies 57* (2019) 223-244.
5. Dubin, J. and Kalsow, G., Participation in Direct Legislation: Evidence from the Voting Booth (1997).
6. Matsusaka, J., Ballot Order Effects in Referendum Elections (2012). [Matsusaka 2012]
7. Mendelsohn, M., Introducing Deliberative Direct Democracy in Canada: Learning from the American Experience, American Review of Canadian Studies, 26:3 (1996) 449-468.
8. Milita, K., Election Laws and Agenda Setting: How Election Law Restrictiveness Shapes the Complexity of State Ballot Measures, *State Politics & Policy Quarterly, Vol. 15, No. 2* (2015) 119-146.

Annex A: Initial Experimental Design

There were two types of information that we intended to provide to the voter in addition to the summary ballot measures: (1) educational materials that elaborate on the practical effects of the measures and (2) lists of supporters and /opponents. In light of research [add citation here] indicating that the position of a ballot measure within the [ballot? Election materials?] is correlated with the degree of support it receives from voters, we also intended to manipulate the position in which our key ballot of interest (*parole* measure) appeared on the ballot.

Figure A1: Variations on supplementary materials regarding the content of the ballot measure

1) None
2) A short *Busy Voter Guide*
3) A longer textual analysis
4) The full text of the measure

As indicated in ***Figure A1***, our design included four levels of educational materials. The first condition represented the 'control': the respondent was given no additional information. The second condition presented the respondent with the *Busy Voter Guide*, a short summary of the ballot measure with easy to read bullet points, pros and cons and and details of what proponents and opponents say about the measure. The third condition presented both the *Busy Voter Guide* and a detailed explanation from the legislative analyst's office providing a thorough analysis of the measure including legislative history. The fourth condition presented the unabridged legal text of the proposed measure.

Figure A2: Variations on supplementary materials regarding support and opposition

1) None
2) List of top financial contributors in support and opposition, with total amount donated
3) List of organizations expressing support or opposition

As indicated in ***Figure A2***, our design also included three levels of supporters/opponents. The first condition represented the 'control': the respondent was given no additional information. The second condition presented the respondent with a list of the top financial *contributors* supporting or opposing the measure including how much they had donated to the cause. The third condition further supplemented the *contributors* information with a list of prominent *endorsers*: organizations that had publicly announced their support for or opposition to the measure.

Finally, as indicated in ***Figure A3***, the design varied the placement of our ballot measure concerning *parole*. Our key item of interest, the *parole* measure, was placed in either the first, second or third position on the ballot.

Figure A3: Sequence of administration of measure of interest (_parole_)

1. Parole	1. XXXX	1. XXXX
2. XXXX	**2. Parole**	2. YYYY
3. YYYY	3. YYYY	**3. Parole**

This ultimately resulted in the 4x3x3 factorial design depicted in ***Table A1*** in which a participant would receive one of 36 possible presentations of our survey.

Given the complexity of the design, we decided that all three ballot measures presented to a respondent would be presented under the same conditions (apart from sequence). If we presented the parole measure with the *Busy Voter Guide* and a list of top financial *contributors*, then the ambulance and clinic measures would be presented in the same way.

In addition to the factorial design, we also wanted to block the participants by political party and education level as discussed above in Experimental Design. We broke the political party into three categories: Republicans, Democrats, and everything else, which included independents and an "other" category that would give participants the opportunity to enter their own party description. We broke the education level into six levels: less than high school, high school degree, some college, college degree, some graduate studies, and graduate degree. This resulted in a total of 18 blocks and within each of those blocks we had 36 potential ballot presentations. We knew that being able to sufficiently power 648 experimental groups with our $500 budget and limited time window would be impossible.

Table A1: *Summary of initial experimental design. The initial experimental design called for 3 treatment conditions related to delivering information about supporters or opponents of a ballot measure, crossed with 4 treatment conditions related to delivering information about the measure itself. These were in turn crossed with 3 conditions related to the ordering of the measure of primary interest, for a total of 36 distinct treatment conditions.*

		Supplementary Information regarding Support or Opposition		
		None	*Contributors*	*Contributors + Endorsers*
Supplementary materials regarding the content of the measure	None	Parole at #1 Parole at #2 Parole at #3	Parole at #1 Parole at #2 Parole at #3	Parole at #1 Parole at #2 Parole at #3
	Busy Voter Guide	Parole at #1 Parole at #2 Parole at #3	Parole at #1 Parole at #2 Parole at #3	Parole at #1 Parole at #2 Parole at #3
	Legislative Analyst's Guide	Parole at #1 Parole at #2 Parole at #3	Parole at #1 Parole at #2 Parole at #3	Parole at #1 Parole at #2 Parole at #3
	Full legal text of ballot measure	Parole at #1 Parole at #2 Parole at #3	Parole at #1 Parole at #2 Parole at #3	Parole at #1 Parole at #2 Parole at #3

Annex B: Demographics Analysis

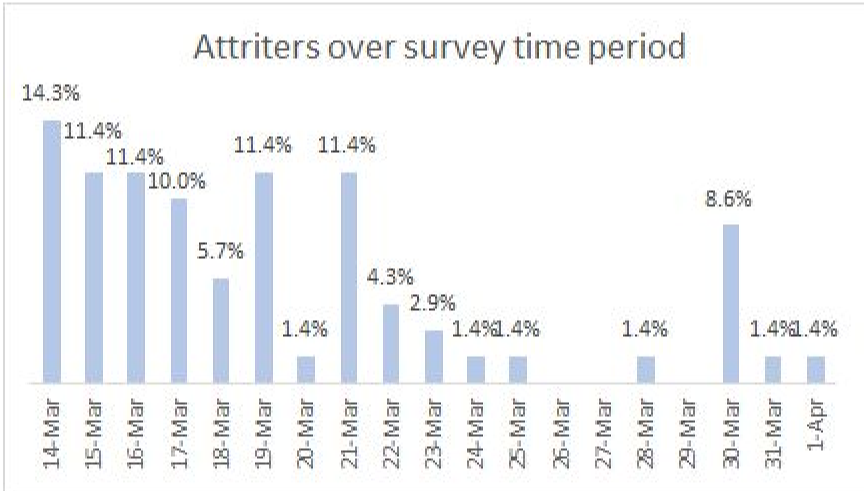

Figure B1: Attrition by day over 4-week survey period.

Table B1: Results for differential attrition check between treatment and control by ballot measure

Treatment vs Control		**Mean control**	**Mean treat**	**df**	**95% CI**	**p-value**
Ballot Measure	Parole	0.1408	0.1639	376	[-0.0946, 0.0484]	0.5261
	Ambulance	0.1221	0.1584	368	[-0.1055, 0.0327]	0.3016
	Clinic	0.1362	0.1694	372	[-0.1049, 0.0384]	0.3621

Table B2: Results for differential attrition check between treatment and control by ballot position

Treatment vs Control		**Mean control**	**Mean treat**	**df**	**95% CI**	**p-value**
Ballot Position	Ballot 1	0.1127	0.1530	365	[-0.1079, 0.0273]	0.2419
	Ballot 2	0.0329	0.0055	291	[0.0010, 0.0538]	0.0419
	Ballot 3	0.0094	0.0219	306	[-0.0375, 0.0125]	0.3271

Table B3: Results for differential attrition check between treatment and control by block

Treatment vs Control		Mean control	Mean treat	df	95% CI	p-value
Block	Democrat + More Education	0.1143	0.2048	153	[-0.1980, 0.01694]	0.09812
	Democrat + Less Education	0.1667	0.1905	141	[-0.2605, 0.2129]	0.8401
	Republican + More Education	0.2222	0.4615	19	[-0.6698, 0.1912]	0.2591
	Republican + Less Education	0.1667	0	5	[-0.2618, 0.5951]	0.3632
	Other + More Education	0.2456	0.1633	104	[-0.0732, 0.2379]	0.2963
	Other + Less Education	0.0833	0.0909	20	[-0.2643, 0.2492]	0.9516

Table B4: Demographic covariate comparison of means across education block, party block and between high/low income earners.

Demographic Covariate		Mean control	Mean treat	df	95% CI	p-value
More education vs Less education	Parole	-31.0677	-20.3857	100	[-18.4594, -2.9045]	0.0076
	Ambulance	-6.4778	-1.7887	108	[-13.5688, 4.1907]	0.2976
	Clinic	21.6008	19.5882	98	[-5.9873, 10.0124]	0.6187
High Income vs Low Income	Parole	-28.1	-29.6987	329	[-4.3694, 7.5669]	0.5986
	Ambulance	-3.7692	-7.4843	328	[-3.3988, 10.8289]	0.305
	Clinic	17.6167	25.3205	331	[-13.7253, -1.6824]	0.01232
Republicans vs Democrats	Parole	-0.5769	-32.6915	28	[17.6010, 46.62812]	9.907e-05
	Ambulance	8.5714	-5.5347	32	[-1.2674, 29.4796]	0.07081
	Clinic	4.68	25.2562	26	[-36.1429, -5.0095]	0.0115