nature human behaviour

Article

https://doi.org/10.1038/s41562-022-01487-4

A 2 million-person, campaign-wide field experiment shows how digital advertising affects voter turnout

Received: 29 November 2021

Accepted: 25 October 2022

Published online: 12 January 2023

Check for updates

Minali Aggarwal[1,10], Jennifer Allen[2,10], Alexander Coppock[1,10], Dan Frankowski[3,10], Solomon Messing[4,10] ✉, Kelly Zhang[5,10], James Barnes[6], Andrew Beasley[7], Harry Hantman[8] & Sylvan Zheng[9]

We present the results of a large, US$8.9 million campaign-wide field experiment, conducted among 2 million moderate- and low-information persuadable voters in five battleground states during the 2020 US presidential election. Treatment group participants were exposed to an 8-month-long advertising programme delivered via social media, designed to persuade people to vote against Donald Trump and for Joe Biden. We found no evidence that the programme increased or decreased turnout on average. We found evidence of differential turnout effects by modelled level of Trump support: the campaign increased voting among Biden leaners by 0.4 percentage points (s.e. = 0.2 pp) and decreased voting among Trump leaners by 0.3 percentage points (s.e. = 0.3 pp) for a difference in conditional average treatment effects of 0.7 points ($t_{1,035,571} = -2.09$; $P = 0.036$; $\widehat{\mathrm{DIC}} = 0.7$ points; 95% confidence interval = −0.014 to 0). An important but exploratory finding is that the strongest differential effects appear in early voting data, which may inform future work on early campaigning in a post-COVID electoral environment. Our results indicate that differential mobilization effects of even large digital advertising campaigns in presidential elections are likely to be modest.

Isolating the causal effects of the billions spent on political advertising in the United States each election cycle has proven to be one of the most difficult research design challenges in the social sciences. Observational studies are vulnerable to confounding; for example, comparisons of vote returns in localities exposed to different levels of advertising suffer from selection bias because advertisements tend to be bought in competitive districts[1,2]. Survey experiments[3–5] address selection bias but typically measure the immediate effects of a single-advertisement dose of advertising. Survey experiments may overstate advertising effects due to exaggerated compliance (attention to an advertisement), experimenter demand effects or unmeasured decay[6–8].

Field experiments often address those measurement concerns by separating the delivery of advertising treatments from the collection of outcomes, either via survey or inspection of electoral returns. While classic work in this area found large but short-lived effects of advertisements in a gubernatorial election[9], a recent meta-analysis of the field experimental evidence concluded that campaign contact has very small effects on vote choice that mostly cannot be distinguished from

[1]Department of Political Science, Yale University, New Haven, CT, USA. [2]MIT Sloan School of Management, Boston, MA, USA. [3]Boxydog, Minneapolis, MN, USA. [4]McCourt School of Public Policy, Georgetown University, Washington, DC, USA. [5]Busara Center for Behavioral Economics, Nairobi, Kenya. [6]Condorsay, New York, NY, USA. [7]Macmillan Publishers, New York, NY, USA. [8]Greenlight, New York, NY, USA. [9]Department of Politics, New York University, New York, NY, USA. [10]These authors contributed equally: Minali Aggarwal, Jennifer Allen, Alexander Coppock, Dan Frankowski, Solomon Messing, Kelly Zhang. ✉e-mail: sm3563@georgetown.edu

https://doi.org/10.1038/s41562-022-01487-4

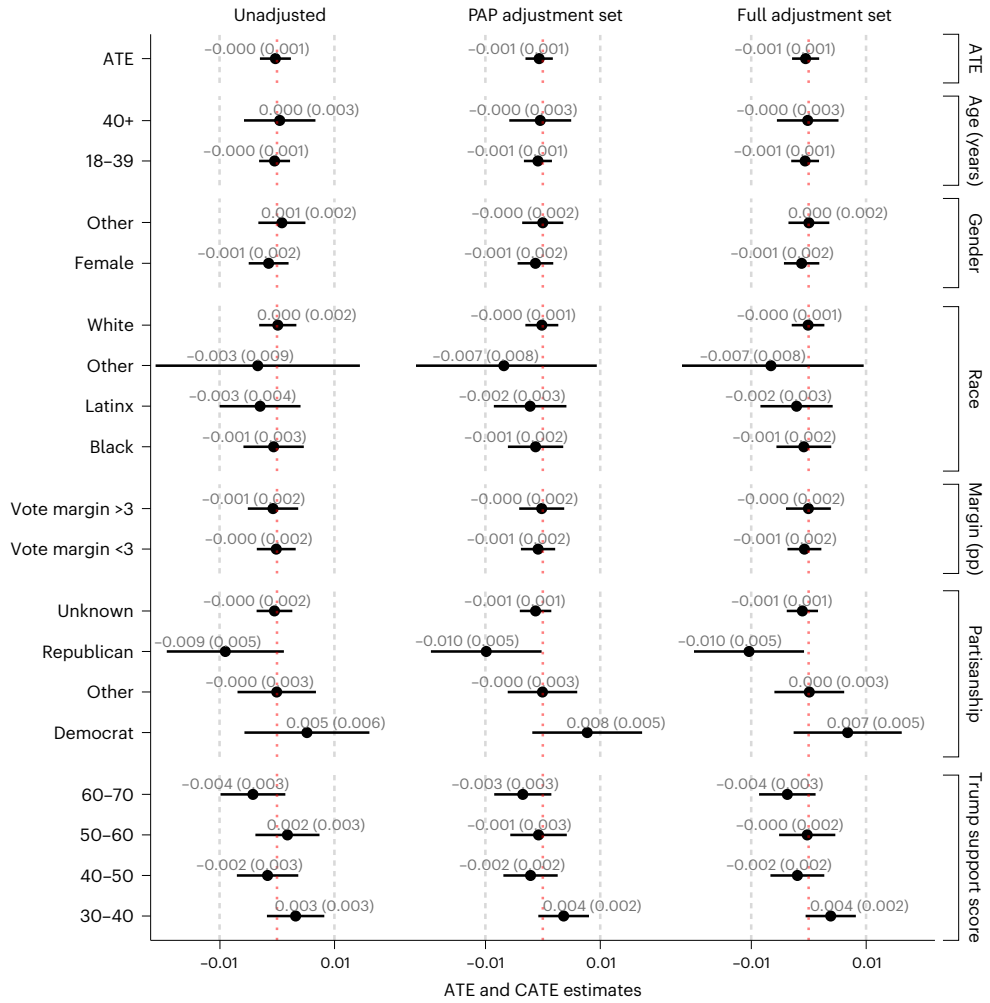

Fig. 1 | Average treatment effects and CATEs of treatment on 2020 turnout under three inverse probability-weighted regression specifications. The three inverse probability-weighted regression specifications were: unadjusted (left column); the PAP adjustment set (middle column); and the full adjustment set (right column). Point estimates and standard errors (in brackets) are reported above each estimate. The error bars represent 95% CIs. Inferential statistics for each estimate are reported in Supplementary Table 1. The numbers were as follows: 1,999,282 (total), 1,379,017 (aged 18–39 years), 620,265 (aged 40+ years), 978,041 (female), 1,021,241 (other gender), 233,546 (Black), 179,036 (Latinx), 1,531,129 (White), 55,571 (other race), 1,337,057 (margin of <3 percentage points), 662,225 (margin of >3 percentage points), 182,945 (democratic partisanship), 71,875 (republican partisanship), 1,442,071 (unknown partisanship), 302,391 (other partisanship), 522,918 (Trump support score of 30–40), 485,371 (Trump support score of 40–50), 478,333 (Trump support score of 50–60) and 512,660 (Trump support score of 60–70).

zero[10]. A persistent worry, however, is that field experiments understate advertising effects because they measure the consequences of only small advertising doses delivered in competitive information environments. A prominent pollster put the critique crisply[11]: 'One group ate a single potato chip, the other had none. Each person was then retested. Would you expect to find that eating a single potato chip affected the health of your subjects?'

Similar challenges arise when studying the questions of whether negative campaigns and political coverage demobilize the electorate. Negative advertisements are hypothesized by some to lower turnout, either because participants are persuaded to abstain rather than vote for a criticized candidate or because they are turned off to politics in general. Using a regression discontinuity design, Spenkuch and Toniatti[12] found that exposure to partisan media increases turnout for in-partisans but decreases turnout for out-partisans. Some laboratory experiments have found that exposure to negative advertisements lowers reported intention to vote, and aggregate analysis of Senate races has found a negative association between tone and turnout[13], although other work has disputed the generalizability of those findings[14] (see also ref. 15 for the authors' reply). Others claim to find the opposite: an observational analysis of Senate campaigns in 1990 found that negative campaigning is associated with higher turnout[16]. Comparing survey respondents who reported that television habits would have exposed them to more or less negative advertising, Freedman and Goldstein[17,18] conclude that negative advertisements stimulate turnout. Still others claim to have found no relationship at all between negative advertising volumes and turnout measured in surveys[19]. More recent observational work has attempted to reconcile these conflicting results by suggesting that demobilization only occurs after a voter decides on a candidate to support, and only when negativity centres on the candidate[20]. Of course, much of this literature describes television advertising rather than digital advertisements, and the collection of studies cited here hardly covers the extensive literature on the effects of negative campaigning on turnout; for a review see ref. 21.

Our design contributes to this literature by enabling us to evaluate the differential mobilization hypothesis at scale with a campaign-level randomized field experiment. In particular, we measure the cumulative impact of an entire US$8.9 million digital advertising campaign aired in

https://doi.org/10.1038/s41562-022-01487-4

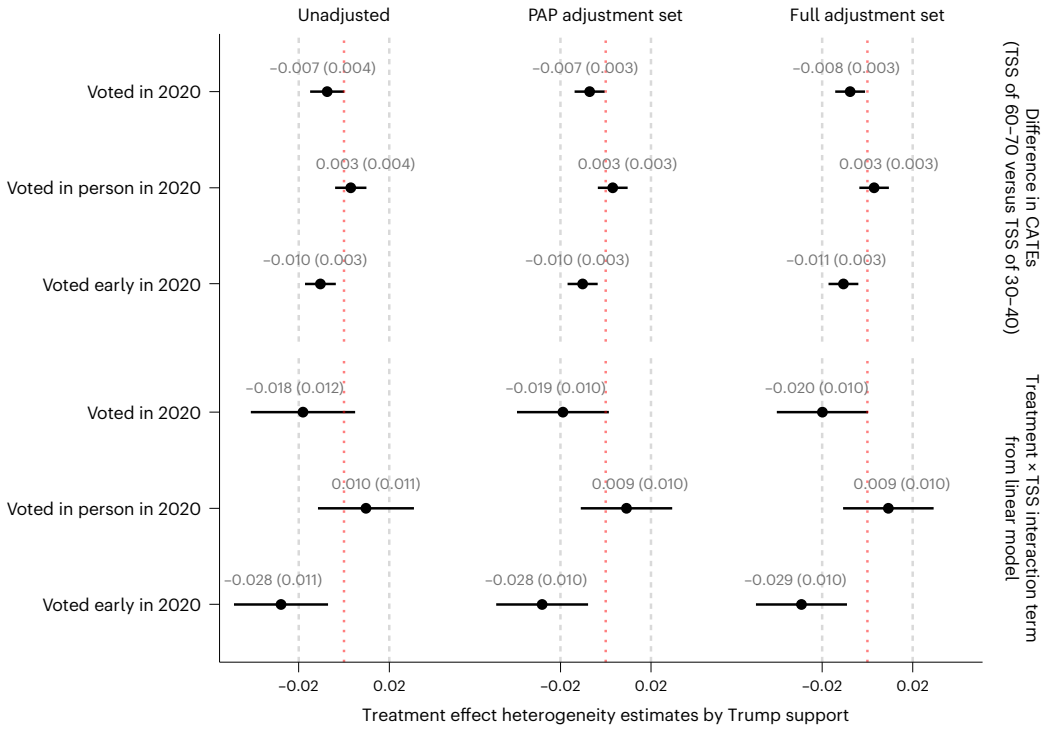

Fig. 2 | Heterogeneous effects of treatment on voting, in-person voting and early voting by Trump support under three inverse probability-weighted regression specifications. The three inverse probability-weighted regression specifications were: unadjusted (left column); the PAP adjustment set (middle column); and the full adjustment set (right column). Point estimates and standard errors (in brackets) are reported above each estimate. The error bars represent 95% CIs. Inferential statistics for each estimate are reported in Supplementary Table 2. The numbers were as follows: 1,035,578 (difference in CATEs estimates) and 1,999,282 (interaction estimates). TSS, Trump support score.

battleground states, with an average of 754 advertisement impressions per treated participant. Our experiment is notable for its sample size and sustained application and dosage of digital advertising.

Our results provide limited support for the differential mobilization hypothesis. According to our pre-registered regression specification to estimate conditional average treatment effects (CATEs), we find that the campaign increased voting among Biden leaners (those with modelled Trump support scores between 30 and 40) by 0.4 percentage points and decreased voting among Trump leaners (those with a Trump support score between 60 and 70) by 0.3 percentage points, resulting in a difference in CATEs (DIC) of 0.7 percentage points ($t_{1,035,571} = -2.09$; $P = 0.036$; $\widehat{\text{DIC}} = 0.7$ points; 95% confidence interval (CI) = −0.014 to 0).

Results

In the 8 months leading up to the 2020 presidential election, Acronym, a prominent left-leaning non-profit organization, conducted a US$8.9 million digital messaging persuasion campaign (see Methods for example advertisements) with the intention of reducing support for Donald Trump and increasing support for Joe Biden in five battleground states: Arizona, Wisconsin, Michigan, North Carolina and Pennsylvania. They kept an experimental holdout, which we use as the control group in this study (see Methods and Table 1 for more details). We compare 2020 turnout according to the voter file maintained by TargetSmart.

We estimate the average causal effects with three weighted least squares regression specifications. The first (unadjusted) is a regression of the outcome on treatment status, with inverse probability weights calculated as the inverse of the probability of each unit being in its observed condition. The second specification (pre-analysis plan (PAP) adjustment set) includes control variables described in the PAP: the Trump support score, the presidential turnout score and a count of whether the participant voted in the 2012, 2016 and 2018 elections. The third specification (full adjustment set) controls for the Trump support score, presidential turnout score, strata fixed effects, indicators for voting in any even-year election between 2000 and 2018 and party membership indicators (Republican, Democrat or unknown relative to other).

We pre-registered that we would estimate CATEs by age, race, gender, party registration and whether Trump's 2016 vote margin in the state was greater than three points. Unfortunately, party registration was not available for a large proportion of the participant pool (72%), so we estimated heterogeneous effects by Trump support score as well. This and all other deviations from the PAP are detailed in the Supplementary Information.

We present the main results of the campaign-level experiment in Fig. 1. Focusing on the pre-registered specification (middle column of facets), we found that the overall effect on turnout (ATE) was −0.06 percentage points, with a robust standard error of 0.12 points ($t_{1,999,277} = -0.52$; $P = 0.60$; $\widehat{\text{ATE}} = -0.0006$; 95% CI = −0.0030 to 0.0017). Using a very narrow equivalence range (plus or minus one-third of a percentage point), we can affirm that our overall estimate is effectively equivalent to zero using the two one-sided tests procedure ($P = 0.013$)[22,23]. We also observed small conditional average effect estimates by age, gender, race and vote margin in 2016 that were not statistically significant.

Turning next to the effects by partisanship and Trump support, we found evidence in favour of the differential mobilization hypothesis. Among those with Trump support scores between 60 and 70, the average effect was a 0.3 point decrease, and among those with scores between 30 and 40, the effect was a 0.4 point increase. As shown in the top row of Fig. 2, the difference in CATEs by Trump support score group was 0.7 percentage points ($t_{1,035,571} = -2.09$; $P = 0.036$; $\widehat{\text{DIC}} = 0.7$ points; 95% CI = −0.014 to 0) according to the PAP adjustment set (see Methods for details).

https://doi.org/10.1038/s41562-022-01487-4

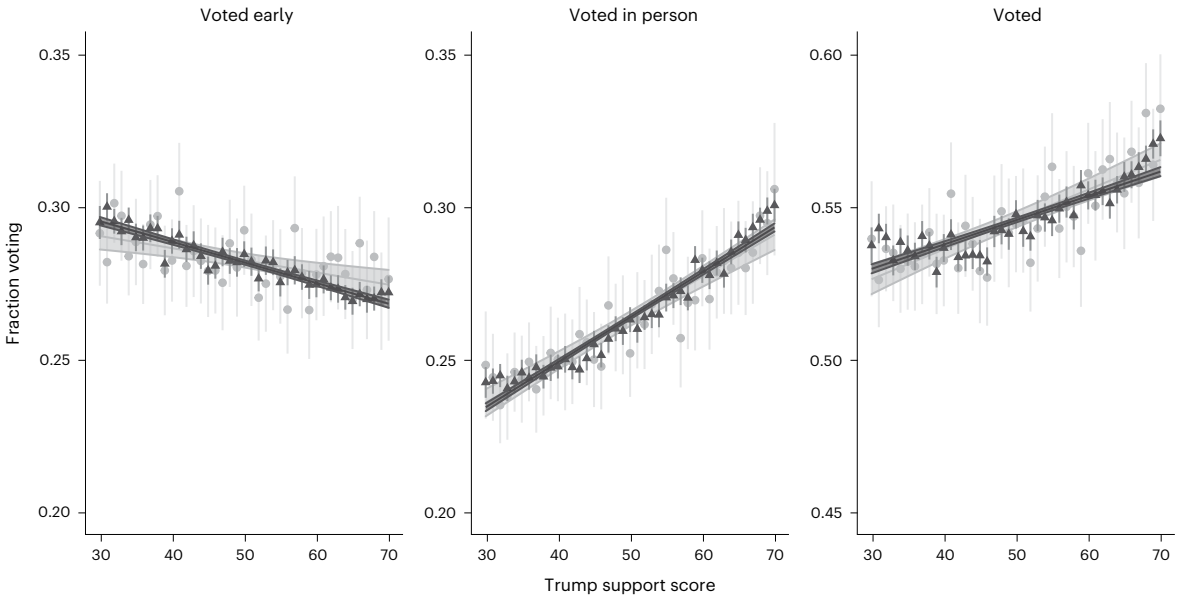

Fig. 3 | 2020 turnout rates by one-point bins of Trump support score and condition. The results are shown for early voting (left), election day voting (middle) and voting regardless of mode (right). The results in black are for the treatment group and the results in grey are for the control group. The error bars represent 95% CIs. Linear predictions from the unadjusted models reported in the bottom left facet of Fig. 2 are overlaid on the binned means, with shaded 95% confidence regions. The vertical scales of all three facets cover a 15 percentage point range, but the ranges differ across facets to emphasize the relevant variation. $n = 1{,}999{,}282$.

Next we examined early and in-person voting in 2020. The next two rows in Fig. 2 report estimates of the difference in CATEs. Here, we see that differential effects of the programme were stronger in our early voting data (1.0 percentage points favouring Biden) than in the in-person voting data (0.3 percentage points favouring Trump) (linear hypothesis test of equality of the differences in CATEs, accounting for the covariance of the estimates: $F_{1,1035571}) = 20.99; P < 0.0001$).

An alternative test using a regression model across the entire range of the Trump support score variable revealed qualitatively similar (if slightly larger) results compared with the difference in CATEs analysis above. The bottom three panels of Fig. 2 report the interaction term from ordinary least squares regressions that linearly interact Trump support with the treatment indicator.

While differential turnout with respect to Trump support was mild, it was strongest for early voting. The scale of the effects can be best appreciated by inspection of Fig. 3. The horizontal axis arrays participants by level of Trump support and the vertical axis shows the average rates of turnout for early voting, in-person voting and all voting. The plots show the CATE estimates within a one-point bin. The estimates themselves are somewhat imprecise owing to the relatively small size of the control group (as reflected in the wider CIs on the untreated means). Overlaid on the CATEs are the predicted average marginal effects from the PAP specification of the interaction term (see the bottom middle facet of Fig. 2). We emphasize that this finding was not pre-registered and should be considered exploratory.

We discuss treatment assignment, balance and example advertisements in the Methods. A description of the treatment assignment process can be found in Fig. 4. Figure 5 shows the differences in pre-treatment covariate means by treatment condition. Figure 6 shows example advertisements used in the manuscript.

Discussion

The experimental literature on the effects of advertising to date has relied on survey experiments that force exposure and measure outcomes immediately, as well as field experiments that deploy doses of treatment that some scholars and practitioners consider to be too small. The challenge in the study of political advertising is to amplify the intensity of treatment to politically meaningful levels in the field while maintaining experimental control. In our view, our design does not suffer from the potato chip critique described in the introduction. In this project, we randomized exposure to the full weight of an 8-month digital advertising campaign, deployed in real time during a contentious election season in battleground states.

What do we learn from this design? First, we provide evidence that persuasion campaigns can indeed cause small differential turnout effects—much smaller than pundits and media commentators often assume, but our field experimental study is large enough to show that these effects are distinct from zero. We found both small mobilizing effects among Biden leaners and small demobilizing effects among Trump leaners. These results shed light on the long-hypothesized causal connection between messages critical of a preferred candidate and decreased turnout.

Second, the strongest differential effects appear in our early voting data, suggesting that it may be increasingly important to advertise early in a post-COVID electoral environment. Because Acronym began its advertising campaign far earlier than other political organizations, our data are particularly well suited to help shed light on this question. We know from previous research that advertisements airing closer to election day tend to have weaker effects, possibly due to saturation and people having already decided how to vote[10]. At the same time, decay effects[9] mean that early advertisements should be expected to have weaker effects over time. Thus, the programme's early advertisements may have had a stronger effect on early voting because they reached voters before they made up their minds, before the media environment became completely saturated and/or before the effects decayed.

Third, this notable dose of advertising had no overall effect on turnout on average. While we tested the full weight of an entire advertising campaign, the fact that this campaign took place in battleground states during a presidential campaign means that exposure to background political advertising and media was relatively high. Acronym's campaign increased that exposure, but not as much as might be expected in a less visible election.

https://doi.org/10.1038/s41562-022-01487-4

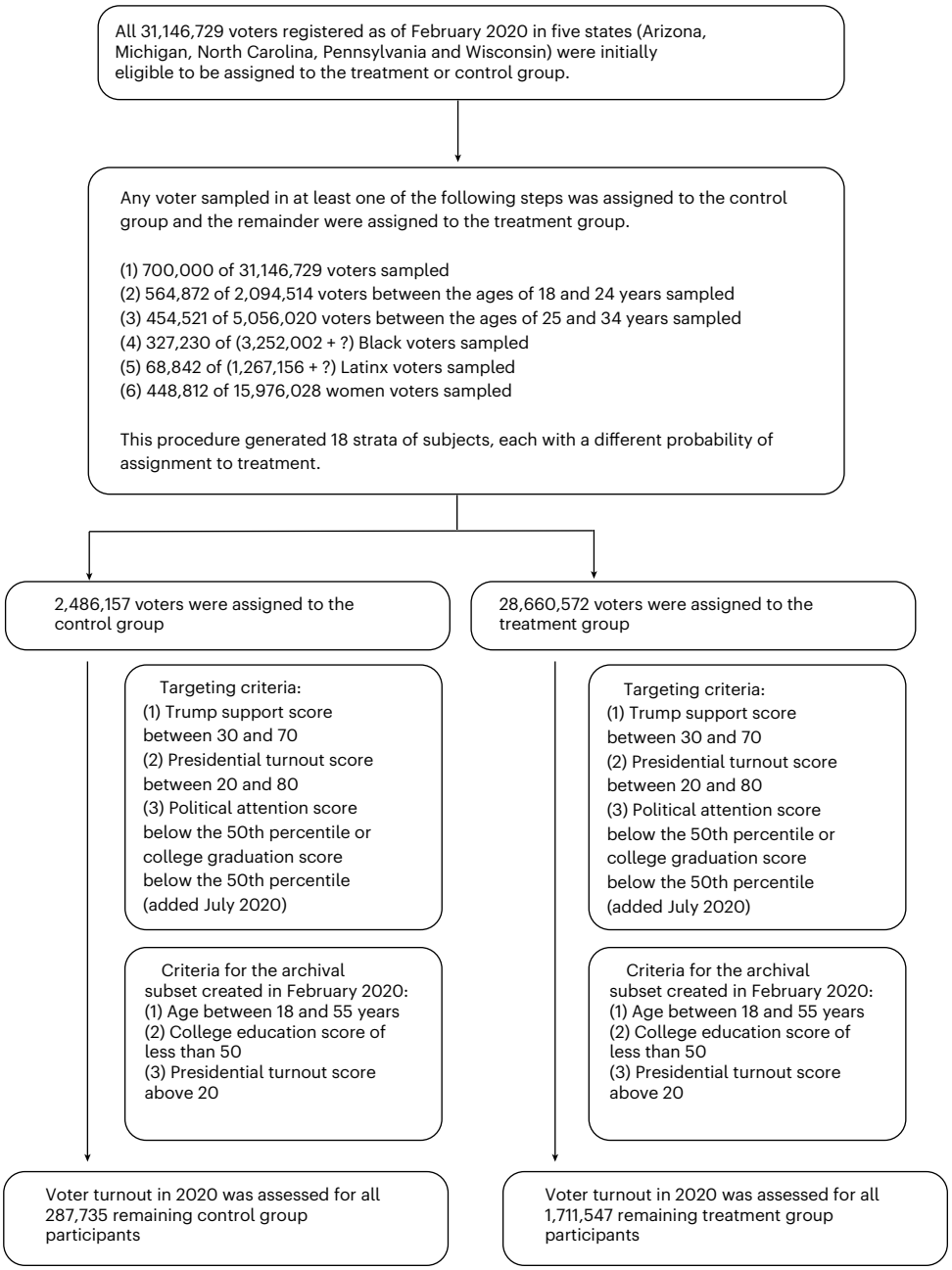

Fig. 4 | Treatment Assignment Flow Chart. Flow chart detailing the experimental sampling, assignment and measurement procedures. From the 31,146,729 eligible voters registered in February 2020 in the states of Arizona, Michigan, North Carolina, Pennsylvania and Wisconsin, a total of 287,735 control group participants and 1,711,547 treatment group participants were assessed for voter turnout. Question marks denote an uncertain number of individuals sampled related to lost imputed race and ethnicity data as described in the "Field experimental design" subsection in Methods below.

We can interpret these small differential turnout effects in two ways. Under one interpretation, the difference reflects persuasion: the advertisements may have lowered evaluations of Trump so much that voters who initially leaned towards voting for him abstained in the end. Under an alternative interpretation, the differential turnout effects might reflect decreased levels of enthusiasm among Trump leaners but increased levels among Biden leaners, without shifting evaluations of Trump. We cannot assess the effects of the treatment on vote choice in this study so our design cannot distinguish between these two channels, although we think either or a combination is plausible.

Zooming out to the broader implications of our study, our results suggest that influencing voter turnout in presidential elections via digital advertisements is expensive. An apples-to-apples comparison with persuasion campaigns in other media is not possible as we lack the requisite information on vote choice. Given our findings, however, the popular narrative that Russia's US$150,000 Facebook advertisement expenditure in 2016 could have caused enough differential turnout to affect the outcome of the election is implausible[24]. This campaign-level field experiment, conducted over 8 months in five battleground states, shows that digital advertisements yielded small returns for presidential campaigns in the 2020 general election. We extrapolate from our study that future digital campaigns of similar scope and size will yield similarly small returns. The quality of this generalization may depend on a number of factors, including how engaging the treatments are (to ensure compliance) and proximity to the election (to mitigate decay).

https://doi.org/10.1038/s41562-022-01487-4

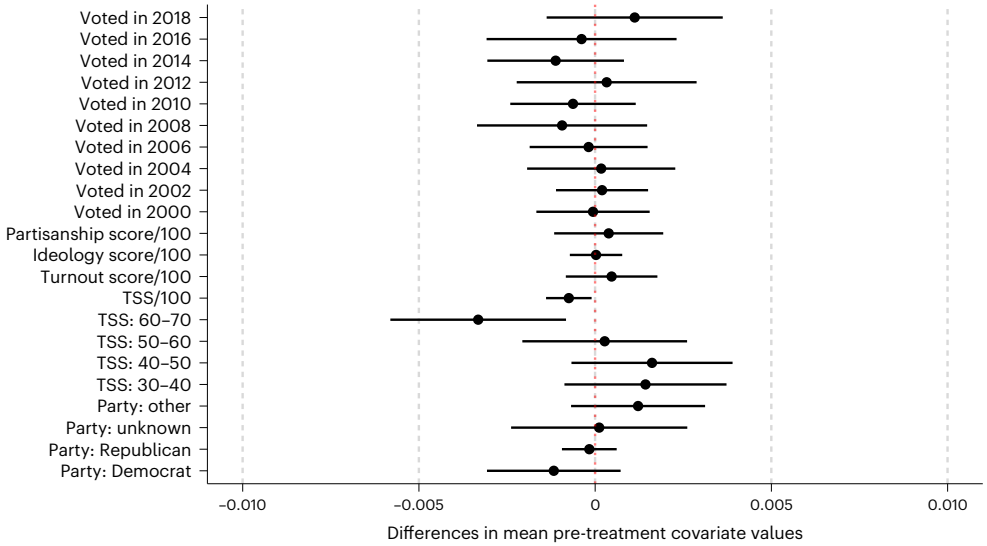

Fig. 5 | Balance on pre-treatment covariates. Point estimates from inverse probability-weighted regressions of the covariate on the treatment indicator are provided. The error bars represent 95% CIs. Inferential statistics are reported in Supplementary Table 3. $n = 1{,}999{,}282$.

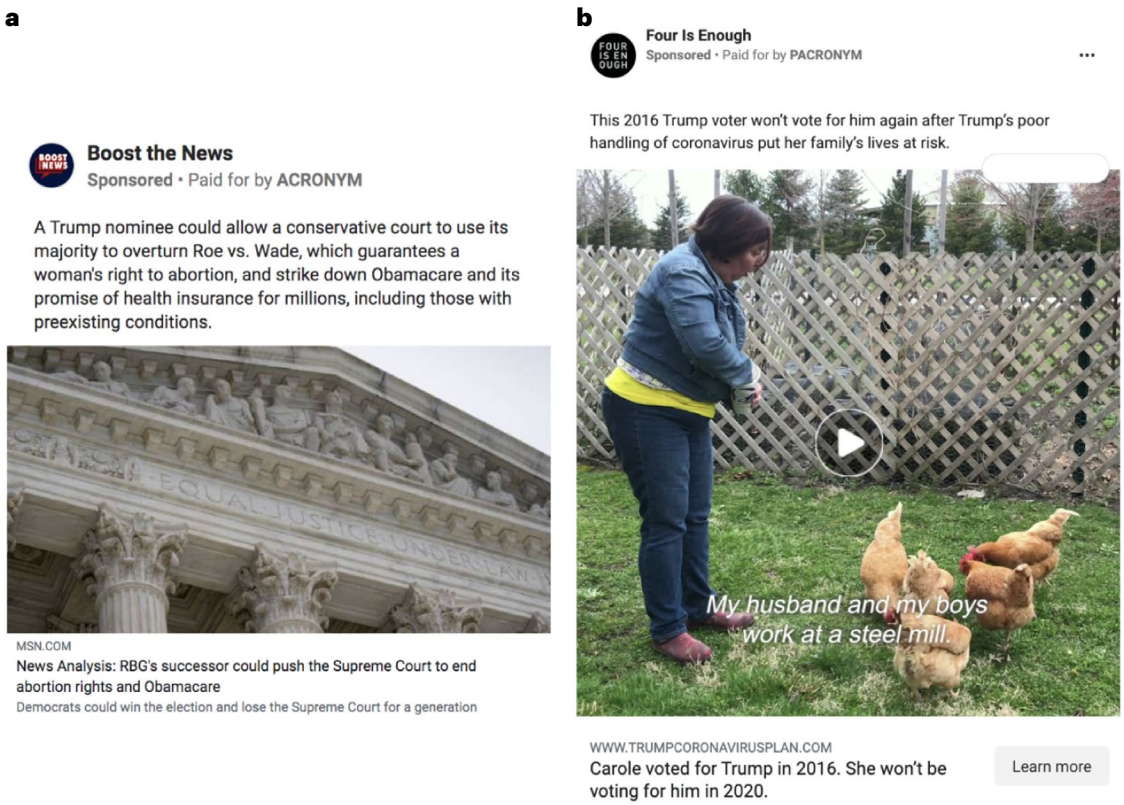

Fig. 6 | Examples of typical advertisement content run in Acronym's persuasion program. a,**b**, Among the types of advertisement content run by Acronym were promoted news stories (**a**) and conventional political video advertisements (**b**).

It is worth noting that this campaign had an emphasis on promoted news stories, alongside conventional political video advertisements. Past survey and observational research has found that emotional appeals may play an important role in political persuasion[25], in part because emotional appeals can be more memorable[26]; however, more recent large-scale experimental evidence has shown the critical importance of information gain in persuasion—in other words, providing a low-information audience with information-rich persuasive content is the most powerful way to influence political behaviour in the aggregate[27].

One reasonable question for our study is how well our findings would generalize to voters who were not eligible for our experimental programme or to other electoral contexts. Conventional wisdom holds that individuals with higher levels of support for one candidate

Table 1 | Experimental strata

Gender	Race	Age (years)	Group size		P_{treat}	2020 voting rate (%)	
			Control	Treatment		Control	Treatment
Female	Black	18–24	4,508	7,605	0.628	48.0	46.4
Female	Black	25–34	3,473	11,952	0.775	42.2	41.3
Female	Black	Other	4,564	26,783	0.854	51.4	51.2
Female	Latinx	18–24	4,831	9,741	0.668	52.2	51.7
Female	Latinx	25–34	3,074	14,743	0.827	46.0	45.0
Female	Latinx	Other	2,727	26,699	0.907	50.7	50.8
Female	Other	18–24	60,999	139,081	0.695	66.0	66.2
Female	Other	25–34	34,406	219,742	0.865	57.0	57.0
Female	Other	Other	19,998	383,115	0.950	62.8	62.5
Other	Black	18–24	14,004	25,313	0.644	36.9	36.9
Other	Black	25–34	11,122	44,850	0.801	27.8	28.1
Other	Black	Other	9,501	69,871	0.880	36.7	36.9
Other	Latinx	18–24	11,145	23,709	0.680	39.3	39.2
Other	Latinx	25–34	5,974	34,290	0.852	33.1	33.1
Other	Latinx	Other	2,835	39,268	0.933	42.6	42.0
Other	Other	18–24	60,726	151,473	0.714	57.9	57.8
Other	Other	25–34	28,067	227,248	0.890	48.3	48.1
Other	Other	Other	5,781	256,064	0.978	57.9	58.6
Total			287,735	1,711,547	0.856	54.6	54.6

or another are harder to persuade and mobilize than those in the middle, suggesting that any turnout effects should be expected to be smaller among those not eligible for the programme. In contrast, the differential mobilization hypothesis holds that we should expect larger demobilization effects among those with stronger attachments to the criticized candidate. Our speculation is that effects would be similarly small at the extremes along the full range of Trump support, although of course we cannot know for sure. Turning to questions of context, it could be that the 2020 election was exceptional because of COVID and the idiosyncrasies of the candidates, so perhaps digital advertising would have larger effects in more typical settings. We can see the logic of this speculation, although our beliefs about larger effects are on the scale of single percentage points, not three or five. Ultimately, learning the answers to these generalizability questions will require further experimentation.

Methods

Research origins, processes and ethics

The experimental design described herein was originally conceived to allow Acronym—a prominent left-leaning 501(c)(4) non-profit organization—to gauge the overall impact of its 'soften the ground persuasion' advertising programme for business reporting purposes. To accomplish this, the organization created a holdout group—a randomly assigned set of people who were not exposed to any of Acronym's advertisements. The research team (including individuals authoring this manuscript) designed and implemented the holdout and helped to administer the holdout group, collect and curate the data and conduct the analysis below.

Yale University's Institutional Review Board reviewed this research and issued a waiver because the data were collected by a third-party, non-academic organization. In the absence of this research, all of the people involved in this study would have received the treatment. The intervention in this case was to remove a random subset of voters from the treatment programme. The main ethical consequence of the research activity was that some participants were not delivered Acronym's advertising but instead were delivered whichever advertisements Facebook and the other advertising platforms might have chosen to show them instead (see the Supplementary Information for a description of the political advertising environment experienced by the control group).

Pre-registration and PAP deviations

The research team pre-registered the design with OSF (https://osf.io/3evfp) in November 2020. We report all of the analyses in that pre-registration. However, because our primary interest is differential turnout, and because we lack party registration information for much of the sample, we submitted an update in December to examine the turnout by Trump support score, which has full coverage (https://osf.io/jkush/). This occurred after seeing early voting data but before seeing final turnout data.

Furthermore, in the pre-registration, we describe one regression specification that includes controls for the Trump support score, the presidential turnout score and a count of vote history. We report this specification in the main text along with two others: an unadjusted specification and a fuller specification that includes the Trump support score, presidential turnout score, strata fixed effects, indicators for voting in any even-year election between 2000 and 2018 and party membership indicators (Republican, Democrat or unknown relative to other).

Subgroups

In the PAP, we specified that we would consider treatment effect heterogeneity by age, race, gender, 2016 vote margin and party registration. We report all of these analyses, but because we lack party registration information for much of the sample, we also include heterogeneity analyses by Trump support score, which is available for all participants. We submitted an update to the registration (after seeing early voting data, but before seeing the final turnout data from the voter file) to use the Trump support score instead of party registration (https://osf.io/jkush/)

Analysis

In the PAP, we describe one regression specification that includes controls for the Trump support score, the presidential turnout score and a count of vote history. We report this specification in the main text along with two others: an unadjusted specification and a fuller specification that includes the Trump support score, presidential turnout score, strata fixed effects, indicators for voting in any even-year election between 2000 and 2018 and party membership indicators (Republican, Democrat or unknown relative to other).

Field experimental design

In the 8 months leading up to the 2020 presidential election, Acronym, a prominent left-leaning non-profit organization, conducted a US$8.9 million digital messaging persuasion campaign with the intention of reducing support for Donald Trump and increasing support for Joe Biden in five battleground states: Arizona, Wisconsin, Michigan, North Carolina and Pennsylvania. The campaign ran paid advertising on Facebook, Instagram and Outbrain advertising networks.

The full experimental design is shown in Fig. 4. In February 2020, eligible participants were randomly assigned to a treatment group that received the messaging programme or to a holdout control group of participants who were never shown any Acronym advertising for the entire 2020 presidential campaign. The random assignment process was unusual and oversampled specific subgroups into the control group. First, a sample from the total population of registered voters was drawn, then successive samples from important subgroups (young people, Black and Latinx voters (where Latinx is a gender-neutral term used as an alternative to Latina or Latino) and women) were drawn with replacement. A voter was assigned to the control group if they were sampled at one or more of these steps. The assignment process resulted in 18 demographic strata, with probabilities of assignment to treatment (P_{treat}) shown in Table 1.

This assignment process differs subtly from standard block random assignment in which fixed numbers of units in each block are assigned to treatment or control. Here, because of the overlapping sample draws, the number of units assigned to treatment in each stratum could differ across realizations of the randomization. This procedure still generates unbiased treatment effect estimates but probably has higher variance than standard block random assignment. Moreover, we are missing race data for 4% of the sample where race was uncoded. At the time of random assignment, Acronym used information from a third-party data provider to infer race categories for these individuals, but these data are no longer available due to the provider's dynamic prediction models. As a result, we cannot calculate the denominators in steps 4 and 5 of the assignment procedure shown in Fig. 4. We therefore estimate the probabilities of assignment from the fraction treated in each group. While these assignment probability estimates are not exact, they are unbiased and quite precise.

Acronym employed specific treatment-targeting criteria to focus on a subset of centrist voters thought to be persuadable. In particular, they directed advertising to voters modelled to have mid-range Trump support and turnout scores. In July of 2020, they narrowed the criteria to exclude voters with above-median political knowledge. We restricted both the treatment and control groups on the basis of all of these criteria.

Due to an oversight during the implementation of the design, we further restricted our analysis to participants between 18 and 55 years old, those with a college education score below 50 and those with a presidential turnout score above 20. After random assignment, the identities of voters assigned to the control group were saved, but the full set of treatment identities were not. Fortunately, an Acronym employee happened to save an exact list of treatment group participants comprising the subset above in February of 2020. By starting from the voter file and filtering on all of the relevant variables, we were able to recover all treatment and control identities in this subset, resulting in an analysis sample of 1,999,282 participants. The reason for this oversight was operational: Acronym planned to deliver advertisements to everyone who met their criteria except those held out to form the control group, so they had no special need to keep a separate list of treatment identities. Over the campaign, the organization made large changes to its data systems. Important data describing the treatment group and treatment strata were lost: most notably, some imputations of race and Hispanic ethnicity. Despite extensive efforts, our attempts to exactly reconstruct these strata and thus the full dataset were unsuccessful, as indicated by imbalance on pre-treatment covariates in the full dataset. Thus, we analyse the archival subset below.

As shown in Fig. 5, the random assignment procedure generated treatment and control groups that were balanced on pre-treatment characteristics in this subset. Out of 21 covariates, only two exhibited statistically significant imbalance: Trump support score ($t_{1,999,280} = -2.260$; $P = 0.024$; $\widehat{ATE} = -0.001$; 95% CI = −0.001 to 0) and Trump support scores between 60 and 70 ($t_{1,999,280} = -2.609$; $P = 0.009$; $\widehat{ATE} = -0.003$; 95% CI = −0.006 to −0.001). Neither of these estimates remained significant after a Benjamini–Hochberg correction[28] to control for the false discovery rate ($P = 0.262$ and 0.200, respectively). Furthermore, when we assessed whether we could predict treatment assignment from the covariates, we found that we could not. An F-test comparing a regression of treatment status on covariates with strata fixed effects versus a restricted model predicting treatment status from strata fixed effects was non-significant ($F_{1999264,-18} = 1.0266$; $P = 0.425$). This omnibus test gave us confidence that within this subset, the experimental design worked as expected. The main way in which the subset differed from the full sample was that it unfortunately excluded voters over 55 years of age.

Advertising campaign and content

The messaging programme consisted of 536 unique paid advertisements. These advertisements largely comprised promoted news—social media posts with links to news articles that were rendered with branding and formatting from the originating news source—and more traditional video and infographic advertisements included before August. Examples of a typical promoted news advertisement and a typical traditional video advertisement can be found in Fig. 6. Acronym produced all of the advertisement spots, conducted audience targeting and purchased all advertisement inventory for the programme. Further examples of treatment stimuli can be found on our OSF site at https://osf.io/ex3kq/.

Using the Facebook Ad Library API, we are also able to calculate the lower bound of spending by Acronym on advertisements containing the words Biden or Trump. Over the experimental period, Acronym spent a minimum of US$368,800 on advertisements containing the word Biden, US$3,254,600 on advertisements containing the word Trump and US$244,500 on advertisements containing both words. Turning to advertisement formats, Acronym spent US$1,275,000 on promoted news advertisements, US$2,288,900 on video advertisements and US$304,000 on other advertising formats (for example, images). While these numbers only represent a minimum spend due to limitations of Facebook's API and further do not include spending on Instagram and Outbrain, they are representative of Acronym's advertising focus. A more detailed description of the content in the programme can be found in Supplementary Fig. 1.

The programme content changed over time. Early in the campaign, the treatment advertisements were a mix of promoted news and traditional video advertisements, but after August 2020, the treatment persuasion programme switched almost entirely to promoted news content. Early in the campaign, the advertisements were mostly anti-Trump, but later advertisements were a mix of anti-Trump and pro-Biden content (Supplementary Figs. 1 and 2).

Using data provided by the Wesleyan Media Project[29], we show that since Acronym's spending started earlier than many other political

actors, its share of the total volume of political advertising on Facebook was higher earlier in the campaign than near election day.

Political activists have long leveraged news to achieve political ends. Recent work has shown just how powerfully the set of news we consume can affect attitudes and beliefs: when a large audience of Fox News viewers were paid to watch CNN for 1 month instead, they responded with less conservative answers to questions about an incoming democratic administration and public health issues[30]. Promoted news in digital political campaigns has been a tactic since at least 2014, when the House GOP created a network of local news domains and promoted them using Google advertisements[31]. In 2018, Well News promoted stories about prominent Blue Dog democrats[32]. The use of promoted news advertisements in political campaigns was common enough in 2020 that Facebook went out of its way to clarify that its political advertisement ban that year (1 week before and in the weeks after the election) applied to its promoted news advertisement product[33].

While there are similarities between promoted news and conventional social media advertisements, one key distinction advertisers often point to is that the messenger in promoted news is a trusted news source rather than a political campaign. For example, Working America ran a boosted news study during the 2020 election and found that promoted news was as effective as traditional advertisement copy, and in fact more effective among Working America union members—a difference they suggest was due to source cues[34].

Dosage and treatment delivery

We gain some appreciation for the dosage of the treatment by considering how participants interacted with the treatment materials. For the promoted news advertisements appearing on Facebook from 5 May through to election day, we found a click-through rate of 1%. The video view rate (which was defined as the fraction of videos that played for at least 2 s while at least halfway on the screen) was 53%. Even if participants did not click the advertisement or watch the video, they were nevertheless exposed to headlines, photographs and accompanying text (see Fig. 6).

As is usually the case with targeted digital campaigns, not all treatment group participants could be successfully identified and served advertisements. Facebook reported that 60% of our treatment group was successfully matched, but did not reveal which units were and were not matched for privacy reasons. In the Supplementary Information, we describe an exploratory analysis that suggests that the 60% matched probably included some false positives. Formally, this implies that our experiment encountered two-sided non-compliance: a large fraction of the assigned treatment group was untreated and a small fraction of the assigned control group was probably treated. We conducted all of our analyses according to the intention-to-treat principle.

One potential concern is that even though the treatment group was exposed to more political advertising than the control group, the control group was nevertheless exposed to some. We think of this issue as a further manifestation of treatment non-compliance. Since political advertising is a small fraction of overall advertising (an estimated 3% of Facebook's Q3 revenue in 2020)[35], exposure in the control group was likely to be small, at least until the final weeks before election day.

The average matched participant received 754 advertisement impressions over the 8 months between March 2020 and election day. In comparison with most field experimental investigations of the effects of political advertisements, this intervention represents a large dose of pro-Biden, anti-Trump information. The full cost of the advertising campaign was US$8.9 million, spread out over a treatment audience of 1,993,216 million (3,322,027 programme-eligible voters × our 60% match rate), amounting to US$4.46 of advertising expenditure per voter.

Statistical software

We implemented all data processing and analysis in R (4.1.1)[36]. For data cleaning, processing and visualization, we used Tidyverse (1.3.1)[37]. For all models, we estimated HC2 robust standard errors, which were implemented using estimatr (0.30.6)[38]. To compare early voting effects with election day effects, we used the linearHypothesis function in the car package (3.1.0)[39].

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

An anonymized replication dataset is available via Dataverse at https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/YMKVA1. TargetSmart generously agreed to make replication data available for this paper. By downloading replication data, researchers agree to use the data only for academic research, agree not to share the data with outside parties and agree not to attempt to re-identify individuals in the dataset in order to download the data.

Code availability

Replication scripts are available via Dataverse at https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/YMKVA1.

References

1. Romer, D., Kenski, K., Winneg, K., Adasiewicz, C. & Jamieson, K. H. *Capturing Campaign Dynamics, 2000 & 2004: The National Annenberg Election Survey* (Univ. Pennsylvania Press, 2006).
2. Sides, J., Vavreck, L. & Warshaw, C. The effect of television advertising in United States elections. *Am. Polit. Sci. Rev.* **116**, 702–718 (2022).
3. Druckman, J. N. The implications of framing effects for citizen competence. *Polit. Behav.* **23**, 225–256 (2001).
4. Druckman, J. N. & Holmes, J. W. Does presidential rhetoric matter? Priming and presidential approval. *Pres. Stud. Q.* **34**, 755–778 (2004).
5. Coppock, A., Hill, S. J. & Vavreck, L. The small effects of political advertising are small regardless of context, message, sender, or receiver: evidence from 59 real-time randomized experiments. *Sci. Adv.* **6**, 36 (2020).
6. Chong, D. & Druckman, J. N. A theory of framing and opinion formation in competitive elite environments. *J. Commun.* **57**, 99–118 (2007).
7. Zaller, J. R. *The Nature and Origins of Mass Opinion* (Cambridge Univ. Press, 1992).
8. Vavreck, L. *The Message Matters: The Economy and Presidential Campaigns* (Princeton Univ. Press, 2009).
9. Gerber, A. S., Gimpel, J. G., Green, D. P. & Shaw, D. R. How large and long-lasting are the persuasive effects of televised campaign ads? Results from a randomized field experiment. *Am. Polit. Sci. Rev.* **105**, 135–150 (2011).
10. Kalla, J. L. & Broockman, D. E. The minimal persuasive effects of campaign contact in general elections: evidence from 49 field experiments. *Am. Polit. Sci. Rev.* **112**, 148–166 (2018).
11. Mellman, M. Are we doing it all wrong? *The Hill* https://thehill.com/opinion/campaign/353720-mark-mellman-are-we-doing-it-all-wrong (2017).
12. Spenkuch, J. L. & Toniatti, D. Political advertising and election results. *Q. J. Econ.* **133**, 1981–2036 (2018).
13. Ansolabehere, S., Iyengar, S., Simon, A. & Valentino, N. Does attack advertising demobilize the electorate? *Am. Polit. Sci. Rev.* **88**, 829–838 (1994).
14. Wattenberg, M. P. & Brians, C. L. Negative campaign advertising: demobilizer or mobilizer?. *Am. Polit. Sci. Rev.* **93**, 891–899 (1999).

15. Ansolabehere, S. D., Iyengar, S. & Simon, A. Replicating experiments using aggregate and survey data: the case of negative advertising and turnout. *Am. Polit. Sci. Rev.* **93**, 901–909 (1999).
16. Kahn, K. F. & Kenney, P. J. Do negative campaigns mobilize or suppress turnout? Clarifying the relationship between negativity and participation. *Am. Polit. Sci. Rev.* **93**, 877–889 (1999).
17. Freedman, P. & Goldstein, K. Measuring media exposure and the effects of negative campaign ads. *Am. J. Polit. Sci.* **43**, 1189–1208 (1999).
18. Goldstein, K. & Freedman, P. Campaign advertising and voter turnout: new evidence for a stimulation effect. *J. Polit.* **64**, 721–740 (2002).
19. Finkel, S. E. & Geer, J. G. A spot check: casting doubt on the demobilizing effect of attack advertising. *Am. J. Polit. Sci.* **42**, 573–595 (1998).
20. Krupnikov, Y. When does negativity demobilize? Tracing the conditional effect of negative campaigning on voter turnout. *Am. J. Polit. Sci.* **55**, 797–813 (2011).
21. Lau, R. R. & Rovner, I. B. Negative campaigning. *Annu. Rev. Polit. Sci.* **12**, 285–306 (2009).
22. Schuirmann, D. J. A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *J. Pharmacokinet. Biopharm.* **15**, 657–680 (1987).
23. Lakens, D. Equivalence tests: a practical primer for *t* tests, correlations, and meta-analyses. *Soc. Psychol. Pers. Sci.* **8**, 355–362 (2017).
24. Samuelsohn, D. Facebook: Russian-linked accounts bought $150,000 in ads during 2016 race. *Politico* https://www.politico.com/story/2017/09/06/facebook-ads-russia-linked-accounts-242401 (2017).
25. Redlawsk, D. *Feeling Politics: Emotion in Political Information Processing* (Springer, 2006).
26. Albertson, B., Dun, L. & Kushner Gadarian, S. in *The Oxford Handbook of Electoral Persuasion* (eds Suhay, E. et al.) 169–183 (Oxford Univ. Press, 2020).
27. Broockman, D. E. & Kalla, J. L. When and why are campaigns' persuasive effects small? Evidence from the 2020 U.S. presidential election. *Am. J. Pol. Sci.* https://doi.org/10.1111/ajps.12724 (2022).
28. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B Methodol.* **57**, 289–300 (1995).
29. Fowler, E.F., Franz, M. & Ridout, T.N. *Political advertising in the United States* (Routledge, 2021).
30. Broockman, D. & Kalla, J. The impacts of selective partisan media exposure: a field experiment with Fox News viewers. Preprint at *OSF* https://doi.org/10.31219/osf.io/jrw26 (2022).
31. House GOP campaign arm created misleading fake news sites. *Talking Points Memo* https://talkingpointsmemo.com/livewire/nrcc-fake-news-sites (2014).
32. Shaw, D. Blue dog-affiliated website blurs line between newsroom and political advertiser. *Sludge* https://readsludge.com/2021/09/17/blue-dog-affiliated-website-blurs-line-between-newsroom-and-political-advertiser/ (2021).
33. Fischer, S. Facebook says new pre-election political ad rules apply to boosted posts. *Axios* https://www.axios.com/facebook-says-new-pre-election-political-ad-rules-apply-to-boosted-posts-ecbab5e5-48c8-42b4-878f-aa7581ed6186.html (2020).
34. 2020: Working America. Fight for a Better America (2020). https://www.fightforbetter.org/workingamerica
35. Levy, A., Rodriguez, S. & Graham, M. Why political campaigns are flooding facebook with ad dollars. *CNBC* https://www.cnbc.com/2020/10/08/trump-biden-pacs-spend-big-on-facebook-as-election-nears.html (2020).
36. R Core Development Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2022).
37. Wickham, H. et al. Welcome to the tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).
38. Blair, G., Cooper, J., Coppock, A., Humphreys, M. & Sonnet, L. estimatr: Fast estimators for design-based inference. R package version 0.30.6 https://CRAN.R-project.org/package=estimatr (2022).
39. Fox, J. & Weisberg, S. *An R Companion to Applied Regression* 3rd edn (Sage, 2019).

Acknowledgements

We thank D. Green, F. Sävje, M. Michelson, B. Sinclair, E. Porter, L. Vavreck, Y. Velez, J. Kalla and D. Broockman for generous early feedback. We also thank E. Franklin Fowler and the Wesleyan Media Project for generously sharing data. We received no specific external funding for this work.

Author contributions

S.M. and J.B. conceived of the initial research. S.M. designed and implemented the original experiment. A.B. and H.H. administered the experiment. D.F., M.A., K.Z., S.Z., J.A. and S.M. contributed to data collection and curation. M.A., J.A. and A.C. analysed the results of the experiment. A.C. and S.M. drafted the initial manuscript and figures. All authors contributed to the revision and editing of the manuscript.

Competing interests

All researchers were employed by Acronym or were contractors thereof during the 2020 election cycle. J.A., S.Z. and H.H. have a substantial financial interest in Facebook. Acronym played a role in conceptualizing and designing the study presented here: all of the authors except A.C. contributed to the design and/or implementation of the experiment while employed by Acronym from January 2020 to January 2021. After the termination of their employment in January 2021, but before working on this manuscript, the other primary authors signed explicit third-party data-sharing agreements on 8 February 2021 to allow data access while they conducted the scientific analysis and reporting presented here, with other institutions. Acronym and the authors agreed in writing to the publication of this manuscript in advance of manuscript analysis and preparation. The authors agreed to provide Acronym with a draft of the manuscript before publication.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41562-022-01487-4.

Correspondence and requests for materials should be addressed to Solomon Messing.

Peer review information *Nature Human Behaviour* thanks the anonymous reviewers for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2023

nature portfolio

Corresponding author(s): Sol Messing

Last updated by author(s): Jul 2, 2022

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	NA
Data analysis	Data analysis relied on R (4.1.1), along with tidyverse (1.3.1), estimatr (0.30.6), car (3.1.0).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The manuscript now includes a data availability statement.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Quantitative field experiment
Research sample	Registered voters in Arizona, Wisconsin, Michigan, North Carolina, and Pennsylvania. between 18 and 55 years old, those with a Presidential Support Score between 30 and 70 (out of 100), a college education score below 50 (out of 100), and a presidential turnout score above 20 (out of 100). Sample includes men, women, and nonbinary/other. Sample was selected to maximize impact of persuasion-ads by the third party originating organization and is very common among persuasion ad targeting campaigns in US elections. Sample is not representative of the U.S. voting population.
Sampling strategy	Stratified sampling and treatment assignment were employed. Overall sample size comprises a census of the target audience. This sampling strategy was employed due to Acronym's attempt to reach the most possible potential voters among their target audience. Target audience decisions were made based on available budget and predicted persuadability of the audience. The "holdout" size was selected to minimize concerns that its size might materially affect the effectiveness of the campaign, based on discussions with Acronym staff and leadership.
Data collection	Voter file data collected by TargetSmart. Aggregated ad exposure data provided by Facebook. Researchers were not blinded to study hypotheses or experimental conditions, however there was no contact between researchers and study participants.
Timing	July-November 2020
Data exclusions	We remove individuals for whom treatment labels were lost and could not be reliably recovered (see Materials and Methods for details).
Non-participation	Some individuals were targeted for ads but may not have logged in to social media, or not paid attention to them. Advertising platforms do not make data available on which people see what ads to protect privacy. It is also possible that some individuals' voting records were not reported accurately in the TargetSmart voter file. Accordingly, estimates in the manuscript are all reported as intent-to-treat estimates, rather that estimates of the treatment effect on the treated.
Randomization	In February of 2020, eligible subjects were randomly assigned to a treatment group that received the messaging program or to a ``hold out'' control group of subjects who were never shown any Acronym advertising for the whole of the 2020 presidential campaign. The random assignment process was unusual due to Acronym's campaign objectives of targeting specific subgroups. First, a sample from the total population of registered voters was drawn, then successive samples from important subgroups (young people, Black and Latinx voters, and women) were drawn with replacement. A voter was assigned to the control group if sampled at one or more of these steps. The assignment process results in 18 demographic strata, each with its own probability of assignment. See paper for additional detail.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging