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

In the format provided by the authors and unedited

Supplementary Information

Contents

A Matching treatment voters to Facebook users **2**

B Description of Ad Content **3**

B.1 Acronym Ad Testing Program 3

B.2 Facebook Ad Library 3

C Analysis of Overall Facebook Ad Environment **5**

D Pre-analysis Plan **9**

D.1 PAP Deviations 14

E Randomization code **15**

F Regression tables **18**

A Matching treatment voters to Facebook users

In order to deliver advertising content to subjects in the treatment group, Acronym uploaded personally identifiable information (PII) from the voter file to Facebook's Custom Audience targeting system. Facebook conducts a "waterfall match," an algorithm that attempts to find Facebook users using the data in the order *provided by the advertiser*, stopping once a match is made. For example, if email is provided first and is sufficient to indicate a match, Facebook does not use secondary, tertiary, or other subsequent fields in the matching process.

To better understand match quality, we leverage the fact that the order in which advertisers provide data fields to Facebook can dramatically change the mechanics of the matching process. While audience match rates were always around 60%, regardless of PII scheme utilized, we hypothesized that prioritization of some types of PII – namely wireless phones and personal emails – would lead to higher match precision. We evaluated match precision by stratifying survey recruitment, targeting by VF age groups, and comparing stratified age group labels to self-reported age from surveys. As we predicted, we found that voter file age and self-reported age were more likely to match when wireless phones and personal email addresses were evaluated earlier in the waterfall matching process. This pattern was especially strong for younger voters. We saw an 82% overall match rate when wireless phones and personal emails were prioritized versus an 81% match rate when name and address were prioritized; for younger voters, the difference was 64% versus 59%. The lack of age data consistency, which ranged from 59 - 82%, suggests at least some degree of false-positive matches. In principle, control units could have be inadvertently received our ads who we did not intend to target.

Formally, this matching process induces two-sided noncompliance: some treatment group units were not matched, so did not receive treatment and some control units could have been matched, so did receive treatment. Our best guess is that being assigned to the treatment group causes a 60 percentage point increase in the probability of being treated, which is important because even with some noncompliance, our assignment to treatment still causes big shifts in who is actually treated. We conduct all analyses according to the intention to treat principle.

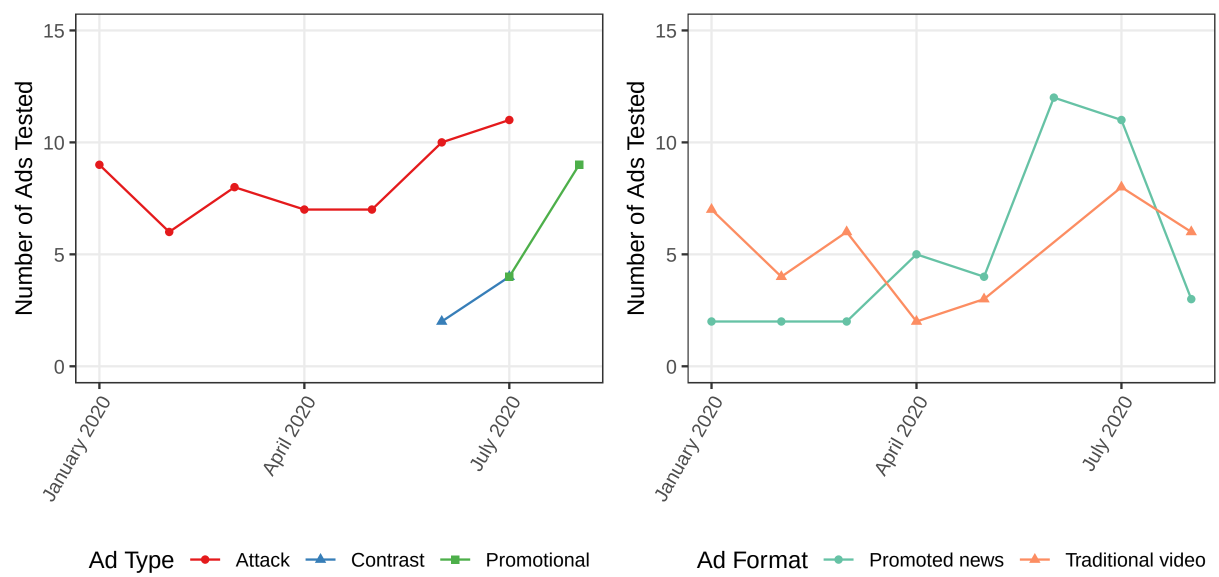
B Description of Ad Content

We describe two separate analyses meant to shed light on the content included in Acronym's advertising program. The analyses show that the program leaned primarily on anti-Trump attack ads, especially early in the year before Joe Biden was the presumptive Democratic nominee, and that Acronym employed a mix of promoted news and traditional video formats, shifting to more news ads closer to the election.

B.1 Acronym Ad Testing Program

Acronym ran in-house message testing of its ads on Facebook and Amazon Mechanical Turk in order to determine which ads were most effective at lowering Trump approval and increasing Biden approval. The ad treatments tested in this program can be found on our OSF site: `https://osf.io/ex3kq/`. These tests informed the allocation of spending in the broader Acronym persuasion program, and, except for a few ads which appeared to inspire backlash in testing, were representative of the overall content of the program. We only include data until August 2020, after which the message testing program focused exclusively on turnout content (which was not shown to the audience included in this paper's analysis).

Figure 1 (left panel) shows the breakdown of ads tested by whether they were anti-Trump attack ads, pro-Biden promotional ads, or contrast ads. As can be seen, most ads tested early in the campaign were anti-Trump attack ads, with more contrast ads and promotional ads tested later in the campaign after Biden became the Democratic nominee. The right panel shows the formats of the ads tested. It shows that Acronym tested more promoted news later in the campaign season.

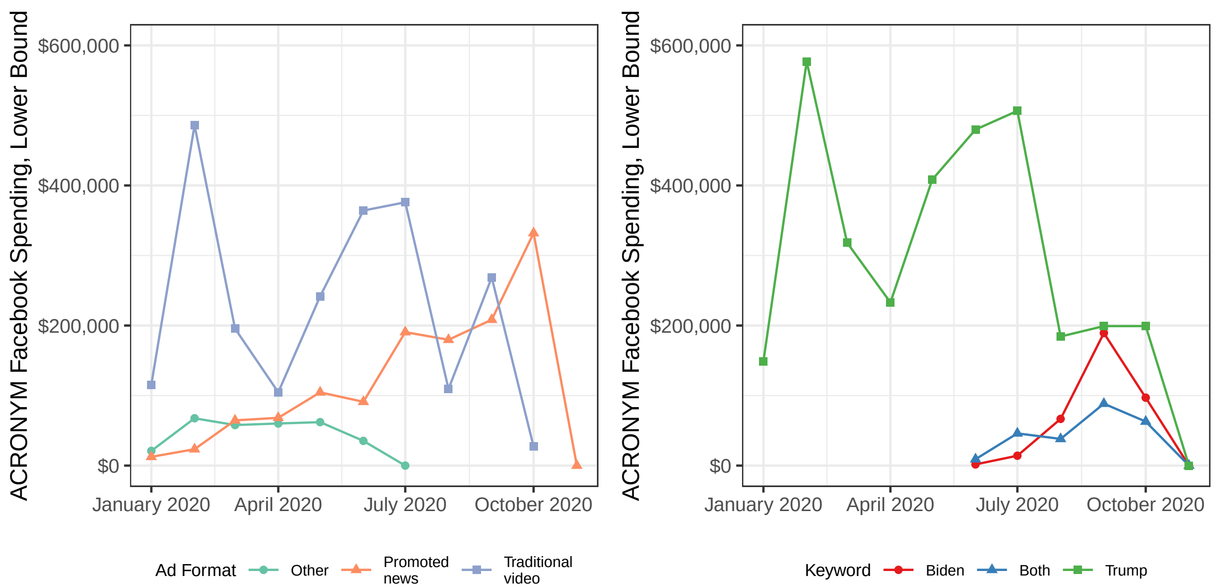


Supplementary Figure 1: Ads tested in Acronym's in-house message testing program, broken down by ad type and format

B.2 Facebook Ad Library

In addition to the in house testing program, we also use Facebook's Ad Archive API, which allows researchers to search political and issue ads run on the platform, to quantify the amount spent by Acronym running ads of different types (see: `https://www.facebook.com/ads/library/api/`. We did this by querying the

API for ads that ran with Acronym's paid-for disclaimer during the time period of the campaign for ads containing the keywords “Biden" or “Trump", and excluding explicit turnout content. While this method allows us to get a rough estimate of the spend of the campaign, there are some drawbacks. First, the API only returns spend information in buckets: $<$100, 100-499, 500-999, 1K-5K, 5K-10K, 10K- 50K, 50K-100K, 100K-200K, 200K-500K, $>$1M. Thus, precise estimates of spend are impossible to calculate, and we only focus on the lower bound of estimates. Second, we only look at ads that contained either “Biden" or “Trump" in the ad, which unfortunately excludes persuasion content that might not contain those keywords (e.g. ads highlighting Kamala Harris). Third, this spending only includes that on Facebook and not on other platforms like Outbrain and Instagram. Despite these limitations, we believe that this data is broadly representative of the spending done by Acronym on its persuasion program. Figure 2 shows the month-by-month spend broken down by keyword and ad format. Early in the campaign, Acronym's ads focused on Trump before including some pro-Biden content later in the campaign. Earlier ads were also mainly traditional videos, while later in the campaign, Acronym shifted its emphasis to promoted news.



Supplementary Figure 2: Lower bound of spending by Acronym on ads containing the words "Biden" or "Trump" on Facebook, over time, by format and keyword

C Analysis of Overall Facebook Ad Environment

We again leverage the Facebook Ad API to get a lower bound of an estimate of all political ad spending in Acronym's targeted states in order to better understand the ad environment on Facebook. To do this, we queried the Facebook Ad API for all ads containing "Biden" or "Trump" that ran from 2/1/2020 - 11/3/2020 on Facebook, where the ads were targeted to at least one of Acronym's program's states: AZ MI, NC, PA, or WI. The lower bound of total spend by all advertisers that meet this criteria is \$349,006,000; unfortunately, we cannot calculate an upper bound since the largest bucket (>\$1 million) is boundless. The results of this analysis can be found in Figure 3.

As can be seen, spending on political ads increases sharply in the leadup to Election Day, with the highest spend coming in October. Note that Facebook prevented new ads from being uploaded after 10/27/2020 (although already existing ad campaigns could continue to run), and froze all political ads after election day, 11/3/2020 [1]. Acronym spent approximately \$3,867,900 to \$5,921,963 on persuasion ads containing the words "Biden" or "Trump" during this period. Comparing both the lower bound of Acronym spend and overall spend as specified above (\$349,006,000), we find that Acronym constituted approximately 1% of spending on Facebook on the presidential campaign in our target states.

An analysis of data shared with us by the Wesleyan Media Project also confirms the case that most spending on Facebook in the 2020 election was heavily concentrated in the last weeks of the campaign [2]. They shared a dataset of pages on Facebook who advertised at least once in a federal election from 9/1/2020 to 11/3/2020 (election day) and their weekly spending on Facebook. The spending totals are much more precise than that given by the Facebook Ad API and cover a larger population of advertisements than the above analysis. According to this data, there was a total of approximately \$977,761,865 spent on federal races on Facebook from 2/1/2020 to 11/3/2020, \$6,603,488 (0.67%) of which was from Acronym. Figure 4 shows the total spending over time, with most of the spending concentrated later in the year (particularly in October).

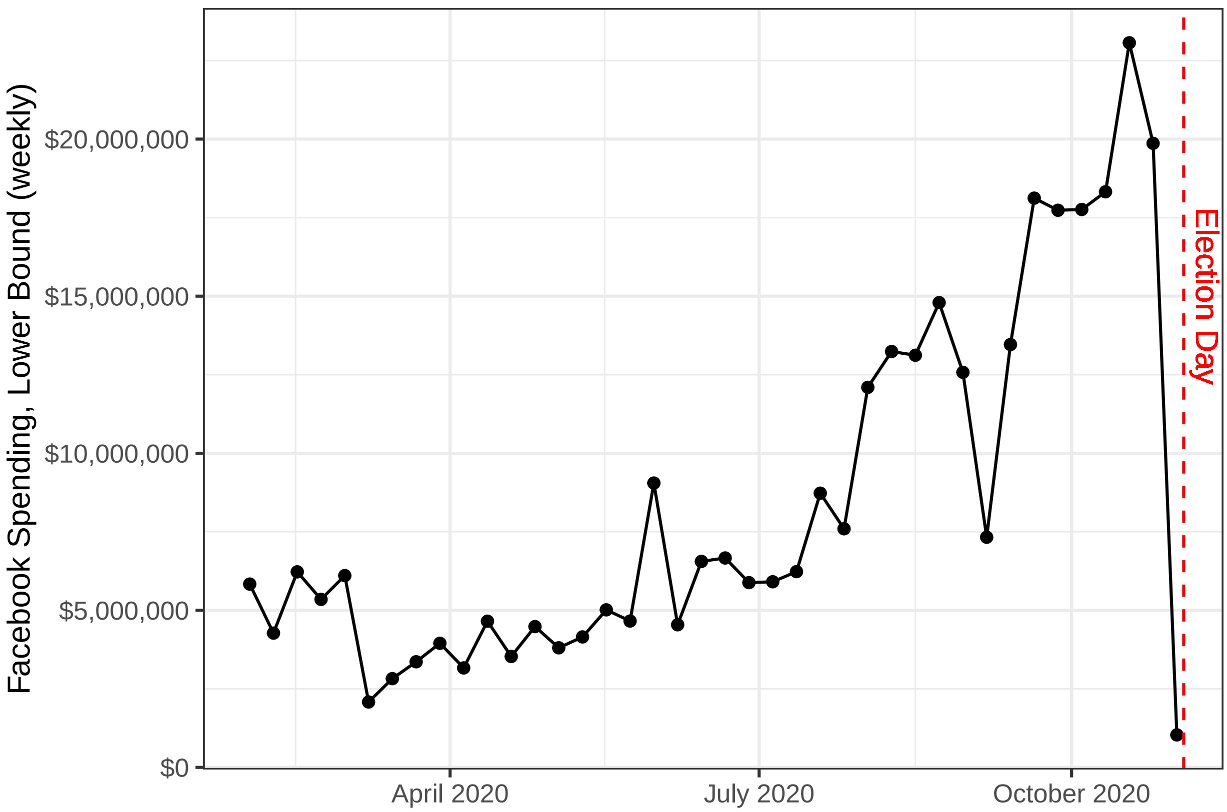
Even though spending by Acronym constituted a relatively low percentage of total spending on political ads on Facebook, we still have reason to believe that Acronym delivered a much higher dose of pro-Biden / anti-Trump persuasion ads to the treatment group than what was seen by the control group. Political advertising is a small fraction of overall advertising (an estimated 3 percent of Facebook's Q3 US ad revenue in 2020) [3], which leads us to infer that the holdout group likely saw a smaller dose of political advertisement than our treatment group, which is the condition we need to hold in order for our experiment to speak to the political science theories under investigation.

A number of other factors make it highly unlikely that that Acronym's holdout audience simply saw a comparable dose of left-leaning political ads, which have to do with the nature of digital advertising. What they saw depends on how ad systems work and what ads other advertising clients were airing. End users of Facebook, Instagram, and Outbrain see a fixed number of ads. The ad served depends on whether the user matches advertiser targeting criteria and if so, the outcome of a second-price auction for the ad slot in question. Advertisers with different targeting criteria will not be in direct competition for ad slots for the same audience. Finally, the modal ad served on Facebook, Instagram, and Outbrain is a commercial ad, for which a clear ROI can be quickly and easily computed to optimize ad bidding, not a political ad, for which the ROI is impossible to compute prior to an election.

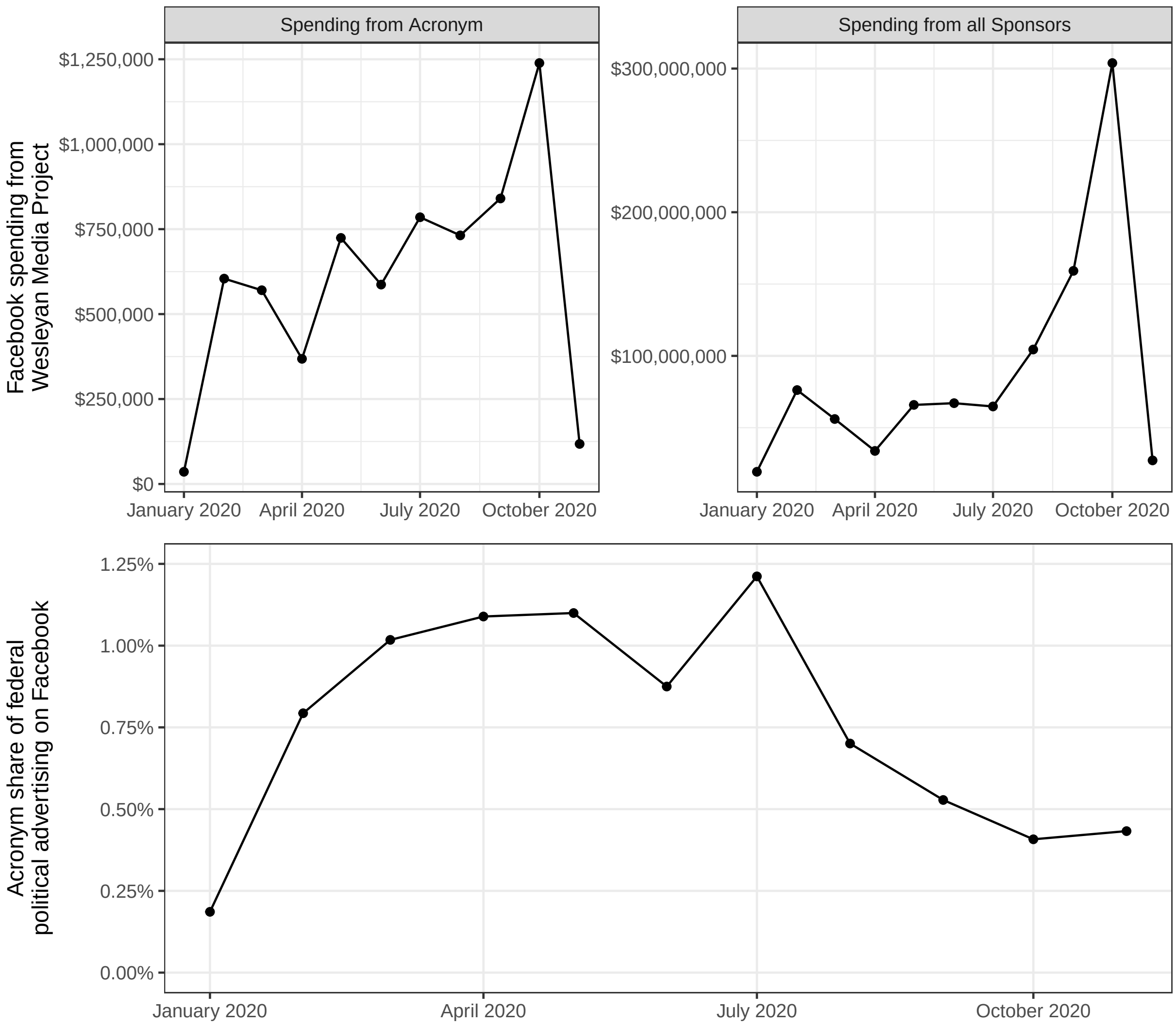
These facts about the Facebook ad environment means that other left-leaning advertisers would need to be targeting the *same audience at the same time with a higher bid than commercial advertisers* for the holdout to contain similar content to the treatment group.

There are a number of reasons this is unlikely. First, traditional campaigns generally target high-education, high-turnout centrist voters for persuasion campaigns, but Acronym targeted an *unconventional audience*, with lower vote-propensity and lower levels of political interest than the conventional "persuasion" audience.

Furthermore, both our analysis and other research shows that campaigns in 2020 spent the majority of funds raised in the two weeks before the election [4], perhaps based on a widespread belief that persuasive effects decay, which has some support in the literature [5].



Supplementary Figure 3: Lower bound of spending by all advertisers on Facebook on ads with the words "Biden" or "Trump" and targeted to at least one of AZ, GA, MI, NC, PA, or WI. Source: Facebook Ad Library



Supplementary Figure 4: Spending on Facebook by pages that advertised in a 2020 federal election. Data was collected and shared by [2]

However, Acronym spent a great deal more ad dollars *earlier* in the campaign than other left-leaning political advertisers in a gamble that 2020 would see unprecedented early voting. In fact, we observe significant early spending (Acronym spent more on persuasion in July than in any other month, see main Figure **??**), and we find much stronger effects on early voting than day of voting (see Figure 3 in the main text). This pattern is consistent with a world in which messaging in the final weeks of the campaign is completely saturated, but ad dollars spent earlier in the campaign were more effectual and did not decay among early voters.

D Pre-analysis Plan

Here we present the original pre-analysis plan registered by Acronym on November 22, 2020, prior to accessing any early vote data at `https://osf.io/3evfp/`.

This PAP covered the analysis plans for two of Acronym's messaging programs (Persuasion and Turnout), but we only study the Persuasion program in this paper. At the moment and we do not plan to write up the turnout experiment due to a number of complications. The decision to allocate resources towards persuasion rather than turnout mean that both the dose and audience were significantly smaller than for the persuasion program (roughly half the budget, half the number of participants, and lower ad-delivery frequency). Additionally, the complete set of IDs included in the treatment group was lost after the employees terminated their employment with Acronym. Combined with the relatively small size of the holdout (control) group, this means that the experiment was not sufficiently large and well-powered to provide good evidence about the impact of campaigns on voter turnout.

In some places, we deviate from the PAP and we note those deviations below.

STG Global Holdout Pre-Analysis Plan

Minali Aggarwal

November 11, 2020

Background

This pre-analysis plan (PAP) builds on our standard operating procedure (SOP) for analysis of "Barometer" experiments running from January to November 2020.

This analysis focuses on the global holdout that we set up in March 2020. Voters in the holdout did not receive ads from our STG Persuasion or Turnout programs. The goal of the final global holdout analysis is to measure the cumulative effects of STG Persuasion and STG Turnout Programs. In particular, we aim to understand whether either program had an effect on verified vote.

STG Persuasion

The STG Persuasion Program ran ads that focused on reducing support for Donald Trump and increasing support for Joe Biden from March to November 2020. We tested many of the ads (and message tracks) in Barometer field experiments and found positive effects on reversed Trump approval and the horse race variable. We also ran 3 global holdout check in surveys throughout the year to measure the program's cumulative effects on our audience, in which we also saw positive effects. We believe that these positive cumulative effects would have an impact on verified vote.

STG Turnout

The STG Turnout Program ran ads that focused on mobilizing left-leaning voters. This program ran from August 2020 to November 3, 2020.

Experimental Design

This experiment falls under the Barometer SOP/PAP for experimental design. Since March, we defined an audience of 5M swing-state voters to target in the STG Persuasion program. We defined a holdout audience that would not receive any ads throughout the rest of the program (450K voters). The rest of the 5M voters were targeted with Acronym ads regularly through November 2020.

We also defined an audience of 1.8M swing-state voters for the STG Turnout program. We defined the holdout audience that would not receive any turnout ads throughout the program (370K voters), while the rest of the audience received ads through November 2020.

To analyze the effect of our program on verified vote, we will compare the holdout audience to the audience that was delivered ads.

Treatment

Treatments for the treatment audience ran from March 2020 to November 2020.

- Cell 0: Control
- Cell 1: Treatment, received STG Persuasion ads

The treatment topics covered the following topic categories:

1. Conservative Messenger
2. Climate
3. Economy
4. Healthcare
5. COVID
6. Racial Justice

Some ads were anti-Trump focused, while others were pro-Biden or Trump/Biden contrast. They also varied by medium: video versus boosted news from sources like Fox News, CNN, Reuters, and more.

Data and Sample

Global Holdout Sampling

- We randomly selected voters from Target Smart Voter File (VF) for the holdout audience, oversampling women, Black, Hispanic, and young people. The remaining voters in AZ, MI, WI, NC, PA, GA were eligible to receive our ads.
- For the STG Persuasion program evaluation, we will filter the audience on the following criteria:
  - Presidential Turnout Score between 20-100
  - TSS between 30-70
  - College Education Score below 50th percentile
  - Political Attention Score below 50th percentile (from Civis)
  - States: AZ, MI, WI, NC, PA
- For the STG Turnout program evaluation, we will filter the audience on the following criteria:
  - Presidential Turnout Score between 0-50
  - TSS between 0-30
  - College Education Score below 50th percentile
  - Political Attention Score below 50th percentile (from Civis)
  - States: AZ, MI, WI, NC, PA, GA

Voter Returns

Because we sampled directly from the Voter File, we can join the Voter Returns by voterbase_id to identify voters and non-voters.

Models

We use the same models as outlined in the Barometer SOP/PAP.

Variables

Outcomes

- Verified Vote, where *verified_vote_any* is:
  - binary indicator for voting in the 2020 presidential election, either by mail or in person
- Verified Mail Vote, where *verified_vote_mail* is:
  - binary indicator for voting by mail in the 2020 presidential election
- Verified In-Person Vote, where *verified_vote_in_person* is:
  - binary indicator for voting in-person in the 2020 presidential election

Subgroups

- Demographic strata
  - Age
    - 18-39
    - 40+
  - Race
    - Black
    - Latinx *Combine Black and Latinx into "Non-white" if Black voters < 10% of sample
    - White
  - Gender, where *is_female*:
    - Binary indicator for whether gender is female
- 2016 Vote Margins, where *vote_margin_over_three_pts*:
  - Binary indicator for whether Trump's vote margin was greater than 3 points in the 2016 election
    - Vote margins > 3% in GA, NC, AZ
    - Vote margins < 3% in PA, WI, MI
- Party registration, where *vb_vf_party* is:
  - 3 categories
    - Republican
    - Democrat
    - Unaffiliated

Controls

- Trump Support Score, where *trump_support_score* is:
  - ts_tsmart_trump_support_score from the VF
- Turnout Score, where *turnout_score* is:
  - ts_tsmart_presidential_general_turnout_score from the VF
- Vote History, where *num_times_voted* is:
  - Number of times voted out of 3 for the 2012, 2016, and 2018 elections

```r
data <- data %>% mutate(
  voted_in_2012 = ifelse(!is.na(vb_vf_g2012), 1, 0),
  voted_in_2016 = ifelse(!is.na(vb_vf_g2016), 1, 0),
  voted_in_2018 = ifelse(!is.na(vb_vf_g2018), 1, 0),
  num_times_voted = voted_in_2012 + voted_in_2016 + voted_in_2018
)
```

For anything not described here, we default to the practices and guidelines outlined in our Barometer SOP/PAP.

D.1 PAP Deviations

Here we detail the deviations from the pre-registration document.

Subgroups In the PAP, we specified that we would consider treatment effect heterogeneity by Age, Race, Gender, 2016 Vote Margin, and Party registration. We do report all those analyses but because we lack party registration information for much of the sample, we also included heterogeneity analyses by Trump support score, which is available for all subjects. We submitted an update to the registration *after* seeing early voting data (but before seeing final turnout data) from the voter file to use Trump Support Score (TSS) instead of party registration (`https://osf.io/jkush/`)

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

Regression Discontinuity Design None of the regression discontinuity analyses were pre-specified.

Archival subset As described in the main text, our final analysis sample differs from the intended sample in the PAP. The reason for this is that when we analyse the full sample, we find clear evidence of experimental imbalance, leading us to believe that something is incorrect in the construction of the full sample. We are able to fully reconstruct the process by which the "archival subset" was created, so we base our inferences on the subset in which we can have confidence. As shown in the main text, balance in this subset is within the normal levels expected in a randomized experiment of this design.

E Randomization code

Here we report the SQL code used to randomly sample units from the voter file in to the holdout control group.

```
drop table if exists "Acronym"."stg_global_holdout_stage1";
drop table if exists "Acronym"."stg_global_holdout_stage2";
create table "Acronym"."stg_global_holdout_stage1" as (
select vb_voterbase_id
    , case when vb_voterbase_race = 'African-American' or (vb_voterbase_race = 'Uncoded' and civis_race
    , case when vb_voterbase_race = 'Hispanic' or (vb_voterbase_race = 'Uncoded' and civis_race = 'HISPA
    , ntl.vb_voterbase_age
    , case when ntl.vb_voterbase_age <18 then 'Under 18'
        when ntl.vb_voterbase_age between 18 and 24 then '18 - 24'
        when ntl.vb_voterbase_age between 25 and 34 then '25 - 34'
        when ntl.vb_voterbase_age between 35 and 44 then '35 - 44'
        when ntl.vb_voterbase_age between 45 and 54 then '45 - 54'
        when ntl.vb_voterbase_age between 55 and 64 then '55 - 64'
        when ntl.vb_voterbase_age > 64 then '65+'
    END as age_category
    , vb_tsmart_city
    , vb_tsmart_state
    , vb_voterbase_gender
    , vb_vf_yob
    , vb_voterbase_dob
    , ntl.vb_tsmart_first_name
    , ntl.vb_tsmart_last_name
    , ntl.vb_tsmart_zip
    , vb_voterbase_phone
    , vb_voterbase_phone_wireless
    , vb_vf_phone
    , email.voterbase_email
    , ts_tsmart_trump_support_score
    , ts_tsmart_presidential_general_turnout_score
    , vb_voterbase_registration_status
from ts.ntl_current ntl
LEFT JOIN tmc.email_current email ON ntl.vb_voterbase_id = email.voterbase_id
where vb_vf_voter_status is not null
    and vb_voterbase_registration_status = 'Registered'
    and vb_voterbase_deceased_flag is null
    and vb_vf_source_state in ('MI','PA','WI','AZ','NC')
    and (vb_voterbase_age >= 18 or vb_voterbase_age is null)
group by 1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20
);
grant all on "Acronym"."stg_global_holdout_stage1" to group Acronym;
select * from "Acronym"."stg_global_holdout_stage1";

drop table if exists "Acronym"."stg_global_holdout_stage2";
create table "Acronym"."global_holdout_stage2" as (
    (select vb_voterbase_id
        , vb_tsmart_first_name
        , vb_tsmart_last_name
```

```
    , vb_tsmart_zip
    , vb_tsmart_city
    , vb_tsmart_state
    , vb_voterbase_age
    , black
    , hispanic
    , vb_voterbase_gender
    , age_category
    , vb_vf_yob
    , vb_voterbase_dob
    , vb_voterbase_phone
    , vb_voterbase_phone_wireless
    , vb_vf_phone
    , voterbase_email
    , ts_tsmart_trump_support_score
    , ts_tsmart_presidential_general_turnout_score
    , vb_voterbase_registration_status
from "Acronym"."stg_global_holdout_stage1"
ORDER BY RANDOM() LIMIT 700000)
--Oversample '18 - 24'
UNION ALL
(select vb_voterbase_id
    , vb_tsmart_first_name
    , vb_tsmart_last_name
    , vb_tsmart_zip
    , vb_tsmart_city
    , vb_tsmart_state
    , vb_voterbase_age
    , black
    , hispanic
    , vb_voterbase_gender
    , age_category
    , vb_vf_yob
    , vb_voterbase_dob
    , vb_voterbase_phone
    , vb_voterbase_phone_wireless
    , vb_vf_phone
    , voterbase_email
    , ts_tsmart_trump_support_score
    , ts_tsmart_presidential_general_turnout_score
    , vb_voterbase_registration_status
from "Acronym"."stg_global_holdout_stage1"
where age_category = '18 - 24'
ORDER BY RANDOM() LIMIT 564872)
UNION ALL
  --Oversample '25 - 34'
  (select vb_voterbase_id
    , vb_tsmart_first_name
    , vb_tsmart_last_name
    , vb_tsmart_zip
    , vb_tsmart_city
    , vb_tsmart_state
    , vb_voterbase_age
    , black
```

```
    , hispanic
    , vb_voterbase_gender
    , age_category
    , vb_vf_yob
    , vb_voterbase_dob
    , vb_voterbase_phone
    , vb_voterbase_phone_wireless
    , vb_vf_phone
    , voterbase_email
    , ts_tsmart_trump_support_score
    , ts_tsmart_presidential_general_turnout_score
    , vb_voterbase_registration_status
  from "Acronym"."stg_global_holdout_stage1"
  where age_category = '25 - 34'
  ORDER BY RANDOM() LIMIT 454521)
UNION ALL
  --Oversample AfAm
  (select vb_voterbase_id
    , vb_tsmart_first_name
    , vb_tsmart_last_name
    , vb_tsmart_zip
    , vb_tsmart_city
    , vb_tsmart_state
    , vb_voterbase_age
    , black
    , hispanic
    , vb_voterbase_gender
    , age_category
    , vb_vf_yob
    , vb_voterbase_dob
    , vb_voterbase_phone
    , vb_voterbase_phone_wireless
    , vb_vf_phone
    , voterbase_email
    , ts_tsmart_trump_support_score
    , ts_tsmart_presidential_general_turnout_score
    , vb_voterbase_registration_status
  from "Acronym"."stg_global_holdout_stage1"
  where black = 'black'
  ORDER BY RANDOM() LIMIT 327230)
UNION ALL
  --Oversample Hispanic American
  (select vb_voterbase_id
    , vb_tsmart_first_name
    , vb_tsmart_last_name
    , vb_tsmart_zip
    , vb_tsmart_city
    , vb_tsmart_state
    , vb_voterbase_age
    , black
    , hispanic
    , vb_voterbase_gender
    , age_category
    , vb_vf_yob
```

```
        , vb_voterbase_dob
        , vb_voterbase_phone
        , vb_voterbase_phone_wireless
        , vb_vf_phone
        , voterbase_email
        , ts_tsmart_trump_support_score
        , ts_tsmart_presidential_general_turnout_score
        , vb_voterbase_registration_status
      from "Acronym"."stg_global_holdout_stage1"
      where hispanic = 'hispanic'
      ORDER BY RANDOM() LIMIT 68842)
    UNION ALL
      --Oversample Women
      (select vb_voterbase_id
        , vb_tsmart_first_name
        , vb_tsmart_last_name
        , vb_tsmart_zip
        , vb_tsmart_city
        , vb_tsmart_state
        , vb_voterbase_age
        , black
        , hispanic
        , vb_voterbase_gender
        , age_category
        , vb_vf_yob
        , vb_voterbase_dob
        , vb_voterbase_phone
        , vb_voterbase_phone_wireless
        , vb_vf_phone
        , voterbase_email
        , ts_tsmart_trump_support_score
        , ts_tsmart_presidential_general_turnout_score
        , vb_voterbase_registration_status
      from "Acronym"."stg_global_holdout_stage1"
      where vb_voterbase_gender = 'Female'
      ORDER BY RANDOM() LIMIT 448812)
);
grant all on "Acronym"."global_holdout_stage2" to group Acronym;
select * from "Acronym"."global_holdout_stage2";
```

F Regression tables

References

[1] What To Know About Facebook Advertising Around the Election. *Meta for Business* URL https://www.facebook.com/business/news/facebook-ads-restriction-2020-us-election.

[2] Neumann, M. *et al.* Identifying and Classifying Federal Election Activity in Online Ads. *Working paper* .

[3] Levy, A. & Rodriguez, S. Why political campaigns are flooding facebook with ad dollars (2020). URL https://www.cnbc.com/2020/10/08/trump-biden-pacs-spend-big-on-facebook-as-election-nears.html.

[4] Ridout, T. N., Fowler, E. F. & Franz, M. M. Spending Fast and Furious: Political Advertising in 2020. *The Forum* **18**, 465–492 (2021). URL `https://www.degruyter.com/document/doi/10.1515/for-2020-2109/html`. Publisher: De Gruyter.

[5] 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. *American Political Science Review* **105**, 135–150 (2011).

Supplementary Table 1: Average and conditional average treatment effects. These estimates correspond to Figure 1 in the main text.

Covariate	Level	Adjustment	Estimate	SE	df	t	p-value	95%CI lower	95%CI upper
ATE	ATE	Unadjusted	-0.000	0.001	1999280	-0.200	0.842	-0.003	0.002
ATE	ATE	PAP adjustment set	-0.001	0.001	1999277	-0.525	0.600	-0.003	0.002
ATE	ATE	Full adjustment set	-0.001	0.001	1999248	-0.436	0.663	-0.003	0.002
Age	18-39	Unadjusted	-0.000	0.001	1379015	-0.304	0.761	-0.003	0.002
Age	18-39	PAP adjustment set	-0.001	0.001	1379012	-0.682	0.495	-0.003	0.002
Age	18-39	Full adjustment set	-0.001	0.001	1378983	-0.493	0.622	-0.003	0.002
Age	40+	Unadjusted	0.000	0.003	620263	0.150	0.881	-0.006	0.007
Age	40+	PAP adjustment set	-0.000	0.003	620260	-0.162	0.871	-0.006	0.005
Age	40+	Full adjustment set	-0.000	0.003	620243	-0.059	0.953	-0.006	0.005
Gender	Female	Unadjusted	-0.001	0.002	978039	-0.821	0.411	-0.005	0.002
Gender	Female	PAP adjustment set	-0.001	0.002	978036	-0.812	0.417	-0.004	0.002
Gender	Female	Full adjustment set	-0.001	0.002	978016	-0.775	0.438	-0.004	0.002
Gender	Other	Unadjusted	0.001	0.002	1021239	0.408	0.683	-0.003	0.005
Gender	Other	PAP adjustment set	-0.000	0.002	1021236	-0.005	0.996	-0.004	0.004
Gender	Other	Full adjustment set	0.000	0.002	1021216	0.025	0.980	-0.004	0.004
Race	Black	Unadjusted	-0.001	0.003	233544	-0.218	0.827	-0.006	0.005
Race	Black	PAP adjustment set	-0.001	0.002	233541	-0.513	0.608	-0.006	0.004
Race	Black	Full adjustment set	-0.001	0.002	233524	-0.342	0.732	-0.006	0.004
Race	Latinx	Unadjusted	-0.003	0.004	179034	-0.818	0.413	-0.010	0.004
Race	Latinx	PAP adjustment set	-0.002	0.003	179031	-0.685	0.493	-0.009	0.004
Race	Latinx	Full adjustment set	-0.002	0.003	179014	-0.655	0.513	-0.008	0.004
Race	Other	Unadjusted	-0.003	0.009	55569	-0.369	0.712	-0.021	0.014
Race	Other	PAP adjustment set	-0.007	0.008	55566	-0.821	0.411	-0.023	0.009
Race	Other	Full adjustment set	-0.007	0.008	55549	-0.797	0.426	-0.023	0.010
Race	White	Unadjusted	0.000	0.002	1531127	0.085	0.932	-0.003	0.003
Race	White	PAP adjustment set	-0.000	0.001	1531124	-0.115	0.908	-0.003	0.003
Race	White	Full adjustment set	-0.000	0.001	1531107	-0.069	0.945	-0.003	0.003
Margin	Vote margin less than 3pp	Unadjusted	-0.000	0.002	1337055	-0.073	0.942	-0.004	0.003
Margin	Vote margin less than 3pp	PAP adjustment set	-0.001	0.002	1337052	-0.543	0.587	-0.004	0.002
Margin	Vote margin less than 3pp	Full adjustment set	-0.001	0.002	1337023	-0.488	0.625	-0.004	0.002
Margin	Vote margin more than 3pp	Unadjusted	-0.001	0.002	662223	-0.310	0.757	-0.005	0.004
Margin	Vote margin more than 3pp	PAP adjustment set	-0.000	0.002	662220	-0.096	0.924	-0.004	0.004
Margin	Vote margin more than 3pp	Full adjustment set	-0.000	0.002	662191	-0.019	0.985	-0.004	0.004
Partisanship	Democrat	Unadjusted	0.005	0.006	182943	0.939	0.348	-0.006	0.016
Partisanship	Democrat	PAP adjustment set	0.008	0.005	182940	1.585	0.113	-0.002	0.017
Partisanship	Democrat	Full adjustment set	0.007	0.005	182914	1.416	0.157	-0.003	0.016
Partisanship	Other	Unadjusted	-0.000	0.003	302389	-0.013	0.990	-0.007	0.007
Partisanship	Other	PAP adjustment set	-0.000	0.003	302386	-0.019	0.985	-0.006	0.006
Partisanship	Other	Full adjustment set	0.000	0.003	302360	0.037	0.970	-0.006	0.006
Partisanship	Republican	Unadjusted	-0.009	0.005	71873	-1.729	0.084	-0.019	0.001
Partisanship	Republican	PAP adjustment set	-0.010	0.005	71870	-2.003	0.045	-0.019	-0.000
Partisanship	Republican	Full adjustment set	-0.010	0.005	71844	-2.123	0.034	-0.020	-0.001
Partisanship	Unknown	Unadjusted	-0.000	0.002	1442069	-0.276	0.783	-0.004	0.003
Partisanship	Unknown	PAP adjustment set	-0.001	0.001	1442066	-0.901	0.367	-0.004	0.001
Partisanship	Unknown	Full adjustment set	-0.001	0.001	1442040	-0.783	0.434	-0.004	0.002
Trump support	30 to 40	Unadjusted	0.003	0.003	522916	1.274	0.203	-0.002	0.008
Trump support	30 to 40	PAP adjustment set	0.004	0.002	522913	1.613	0.107	-0.001	0.008
Trump support	30 to 40	Full adjustment set	0.004	0.002	522884	1.724	0.085	-0.001	0.008
Trump support	40 to 50	Unadjusted	-0.002	0.003	485369	-0.601	0.548	-0.007	0.004
Trump support	40 to 50	PAP adjustment set	-0.002	0.002	485366	-0.885	0.376	-0.007	0.003
Trump support	40 to 50	Full adjustment set	-0.002	0.002	485337	-0.818	0.413	-0.007	0.003
Trump support	50 to 60	Unadjusted	0.002	0.003	478331	0.640	0.522	-0.004	0.007
Trump support	50 to 60	PAP adjustment set	-0.001	0.003	478328	-0.296	0.767	-0.006	0.004
Trump support	50 to 60	Full adjustment set	-0.000	0.002	478299	-0.098	0.922	-0.005	0.005
Trump support	60 to 70	Unadjusted	-0.004	0.003	512658	-1.454	0.146	-0.010	0.001
Trump support	60 to 70	PAP adjustment set	-0.003	0.003	512655	-1.373	0.170	-0.008	0.001
Trump support	60 to 70	Full adjustment set	-0.004	0.003	512626	-1.477	0.140	-0.009	0.001

Supplementary Table 2: The heterogeneous effects of treatment by Trump support. These estimates correspond to Figure 2 in the main text.

Target	Outcome	Adjustment	Estimate	SE	df	t	p-value	95%CI lower	95%CI upper
Difference-in-CATEs TSS 60-70 versus TSS 30-40	Voted early in 2020	Unadjusted	-0.010	0.003	1035574	-3.002	0.003	-0.017	-0.004
Difference-in-CATEs TSS 60-70 versus TSS 30-40	Voted early in 2020	PAP adjustment set	-0.010	0.003	1035571	-3.005	0.003	-0.017	-0.004
Difference-in-CATEs TSS 60-70 versus TSS 30-40	Voted early in 2020	Full adjustment set	-0.011	0.003	1035542	-3.151	0.002	-0.017	-0.004
Difference-in-CATEs TSS 60-70 versus TSS 30-40	Voted in person in 2020	Unadjusted	0.003	0.004	1035574	0.847	0.397	-0.004	0.010
Difference-in-CATEs TSS 60-70 versus TSS 30-40	Voted in person in 2020	PAP adjustment set	0.003	0.003	1035571	0.922	0.356	-0.003	0.010
Difference-in-CATEs TSS 60-70 versus TSS 30-40	Voted in person in 2020	Full adjustment set	0.003	0.003	1035542	0.890	0.374	-0.004	0.009
Difference-in-CATEs TSS 60-70 versus TSS 30-40	Voted in 2020	Unadjusted	-0.007	0.004	1035574	-1.933	0.053	-0.015	0.000
Difference-in-CATEs TSS 60-70 versus TSS 30-40	Voted in 2020	PAP adjustment set	-0.007	0.003	1035571	-2.093	0.036	-0.014	-0.000
Difference-in-CATEs TSS 60-70 versus TSS 30-40	Voted in 2020	Full adjustment set	-0.008	0.003	1035542	-2.268	0.023	-0.014	-0.001
Treatment*TSS interaction term from linear model	Voted early in 2020	Unadjusted	-0.028	0.011	1999278	-2.624	0.009	-0.049	-0.007
Treatment*TSS interaction term from linear model	Voted early in 2020	PAP adjustment set	-0.028	0.010	1999276	-2.712	0.007	-0.048	-0.008
Treatment*TSS interaction term from linear model	Voted early in 2020	Full adjustment set	-0.029	0.010	1999247	-2.840	0.005	-0.049	-0.009
Treatment*TSS interaction term from linear model	Voted in person in 2020	Unadjusted	0.010	0.011	1999278	0.897	0.370	-0.012	0.031
Treatment*TSS interaction term from linear model	Voted in person in 2020	PAP adjustment set	0.009	0.010	1999276	0.892	0.373	-0.011	0.029
Treatment*TSS interaction term from linear model	Voted in person in 2020	Full adjustment set	0.009	0.010	1999247	0.907	0.365	-0.011	0.029
Treatment*TSS interaction term from linear model	Voted in 2020	Unadjusted	-0.018	0.012	1999278	-1.542	0.123	-0.041	0.005
Treatment*TSS interaction term from linear model	Voted in 2020	PAP adjustment set	-0.019	0.010	1999276	-1.826	0.068	-0.039	0.001
Treatment*TSS interaction term from linear model	Voted in 2020	Full adjustment set	-0.020	0.010	1999247	-1.935	0.053	-0.040	0.000

Supplementary Table 3: Balance estimates. These estimates correspond to Figure 5 in the main text.

Covariate	Estimate	SE	df	t	p-value	p-value (BH correction)	95%CI lower	95%CI upper
Party: Democrat	-0.001	0.001	1999280	-1.211	0.226	0.810	-0.003	0.001
Party: Republican	-0.000	0.000	1999280	-0.414	0.679	0.944	-0.001	0.001
Party: Unknown	0.000	0.001	1999280	0.091	0.927	0.944	-0.002	0.003
Party: Other	0.001	0.001	1999280	1.257	0.209	0.810	-0.001	0.003
TSS: 30-40	0.001	0.001	1999280	1.219	0.223	0.810	-0.001	0.004
TSS: 40-50	0.002	0.001	1999280	1.385	0.166	0.810	-0.001	0.004
TSS: 50-60	0.000	0.001	1999280	0.229	0.819	0.944	-0.002	0.003
TSS: 60-70	-0.003	0.001	1999280	-2.609	0.009	0.200	-0.006	-0.001
Trump support score / 100	-0.001	0.000	1999280	-2.260	0.024	0.262	-0.001	-0.000
Turnout score / 100	0.000	0.001	1999280	0.707	0.480	0.944	-0.001	0.002
Ideology score / 100	0.000	0.000	1999280	0.070	0.944	0.944	-0.001	0.001
Partisanship score / 100	0.000	0.001	1999280	0.489	0.625	0.944	-0.001	0.002
Voted in 2000	-0.000	0.001	1999280	-0.072	0.943	0.944	-0.002	0.002
Voted in 2002	0.000	0.001	1999280	0.296	0.767	0.944	-0.001	0.002
Voted in 2004	0.000	0.001	1999280	0.161	0.872	0.944	-0.002	0.002
Voted in 2006	-0.000	0.001	1999280	-0.215	0.829	0.944	-0.002	0.001
Voted in 2008	-0.001	0.001	1999280	-0.761	0.447	0.944	-0.003	0.001
Voted in 2010	-0.001	0.001	1999280	-0.690	0.490	0.944	-0.002	0.001
Voted in 2012	0.000	0.001	1999280	0.252	0.801	0.944	-0.002	0.003
Voted in 2014	-0.001	0.001	1999280	-1.132	0.258	0.810	-0.003	0.001
Voted in 2016	-0.000	0.001	1999280	-0.279	0.780	0.944	-0.003	0.002
Voted in 2018	0.001	0.001	1999280	0.881	0.378	0.944	-0.001	0.004