Supplementary Information to *Does Micro-targeting Work*?

Musashi Jacobs-Haruakwa

Contents

Word Count: 8972 words (excluding table of contents)

Appendix A: Design and Implementation

Design Summary

Each of the respondents $i \in N$ was allocated one of five advertisements $D_{i,a}$, $a \in [1,5]$. The five advertisements, detailed in Table 1 of the main article, were selected from the set of all anti-Biden advertisements run by the Donald Trump campaign on Facebook and YouTube. The criteria for selection are: all advertisements are clearly tailored to different audiences, but no single advertisement is likely to out-perform all others. Advertisements focusing on Joseph Biden were chosen in favor of Donald Trump because two weeks prior to the election it was uncertain would Trump would drop out due to COVID-19. Although Coppock, Hill, and Vavreck (2020) find no evidence of asymmetric effects for attack versus promotional advertisements, attack advertisements were chosen because there is an extensive literature focusing on the normative issues surrounding negative advertising (Ansolabehere and Iyengar 1995).

In the first stage, respondents first answered demographic and political questions regarding age, gender, race, income group, state, interest in news, whether they thought the country was on the right track for the past four years, a seven-point partisan identification scale, and a five-point liberal-conservative ideological self-identification scale. The format and wording of these items was adopted from the American National Election Study.

Respondents were then given a prompt stating that they would be shown a short advertisement, and then served one of the five advertisements detailed above at random. Randomization was done using permuted block randomization over a discrete uniform distribution; $P(a) \sim \mathcal{U}\{1,5\}$.

After viewing the advertisement, respondents were asked three post-treatment items (plus a manipulation check). These included their favorability rating of Biden and Trump on one to five scale, and their voting intention. Given that the survey was run very close to the actual election and there were high rates of early voting, respondents were given the option to state whom they had already voted for.

The data gathered in the first stage was then used to train a predictive model to learn Biden favorability as a function of the pre-treatment covariates and advertisement shown. Thus given any profile of pre-treatment covariates, the predicted outcome under each of the five advertisements would be calculated $\{\hat{Y}(D_1), \hat{Y}(D_2), ...\hat{Y}(D_5)\}$, and the advertisement that minimized Biden favorability, $D^*(X_i)$, would be allocated:

$$D^*(X_i) := argmin_a \ f(X_i, D_{i,a})$$

In the second stage, respondents first answered the same pre-treatment questions. These answers were sent to a server-side `Python` kernel, which given the pre-treatment covariates X_i, used the pre-fitted RF model to allocate the optimal advertisement $D^*(X_i)$. The respondent then watched the advertisement and answered the same three post-treatment items. The first and second stage were therefore indistinguishable from the perspective of the respondent, as they were unaware which advertisements they were not shown and how the advertisement was allocated to them until an end-of-survey debrief.

Provided that there are no systematic differences between the first and second stage, we can compare the outcome between randomly assigned stage 1, $\mathbb{E}_a[\mathbb{E}_i[Y_i(D_{i,a})]]$, and optimally assigned stage 2, $\mathbb{E}_i[Y_i(D^*(X_i))]$, as an estimator of the average effect of targeting:

Stated in potential outcomes notation:[1]

$$ATE = \mathbb{E}_i[Y_i(D^*(X_i))] - \mathbb{E}_a[\mathbb{E}_i[Y_i(D_{i,a})]]$$

[1] *Note on Notation:* In the control group (right-hand term of equation), I average the value of the outcome over all individuals ($\mathbb{E}_i$) and an equal probability of seeing any particular advertisement ($\mathbb{E}_a$). In the treatment group (left-hand term of equation), the advertisement allocation is not random, but a function of pre-determined respondent traits ($D^*(X_i)$).

Following a power analysis based on a simulated run of the experiment using the replication data for Coppock, Hill, and Vavreck (2020), the total N was set at $2,400$,[2] with $1,500$ respondents allocated to the control group and 900 respondents allocated to the treatment group. All participants are United States citizens, resident in the United States, of voting age. Because the design requires a sizable sample to undertake the experiment in a short window of time, the experiment was conducted online rather than in-person or over the phone. Respondents were recruited via the survey provider Prolific and redirected to a custom-built website at `https://survey.<REDACTED>.org`.

Case and Advertisement Selection

The 2020 US presidential election was chosen as the setting for this experiment for several reasons. Enormous campaign spending (Baldwin-Philippi 2017) and a relative lack of regulation (Dobber, Ó Fathaigh, and Zuiderveen Borgesius 2019) made the US a relevant setting with a wide selection of advertisements made for a targeted campaign.

The five advertisements were selected from nearly one hundred videos with a length between 15 and 35 seconds posted by the Trump campaign to their YouTube channel in the final five months of the 2020 United States presidential contest. These were downloaded using the `youtube-dl` tool and hosted on the survey website listed above.

After narrowing down the videos to 10 likely candidates, I asked a panel of ten doctoral students at the our university to help select five advertisements based on the criteria that:

- The advertisements were clearly tailored to different audiences.
- No one advertisement was likely to outperform all others for all respondents.

Irregularities and Attention Check

Responses that failed one of two checks have been omitted from the data used for this article. Prolific provides basic demographic data on respondents that can be downloaded after respondents have completed the survey. Responses where there were considerable discrepancies between answers and supplied demographic information were rejected.

There was also a attention check immediately after the advertisement, which asked respondents which campaign ran the advertisement ("My name is ______ and I approve of this message"). Given that the answer was provided in the last few seconds of the advertisement, and the question was asked less than a few seconds later, I assumed that respondents who failed this were not paying attention to the video and therefore rejected their responses from the final data.

Sample Representativeness

As mentioned, due to resource constraints it was not possible to procure a representative sample. Nevertheless, the sample provided by Prolific covered all 50 states and Washington DC, with a roughly proportional number of respondents to the population of each state (see table A.3).

Table A.1: Race and Ethnic Representation

Race	Census (2019)	Sample
White	60.1%	66.29%
Black	13.4%	9.16%
Hispanic	18.5%	7.03%
Asian	5.9%	12.8%

[2]See Figure A.4.

Table A.2: Income Bound Representation

Income Bound	Census (2019)	Sample
\$0 to \$53.5k	40%	54.67%
\$53.6k to \$109.7k	30%	29.4%

Table A.3: Number of respondents per state (or district), with percentage over/under representation.

State	N	Distortion	State	N	Distortion
Alabama	17	0.74%	Montana	5	0.1%
Alaska	3	0.09%	Nebraska	12	0.06%
Arizona	53	-0.13%	Nevada	31	-0.43%
Arkansas	10	0.48%	New Hampshire	6	0.15%
California	307	-1.54%	New Jersey	79	-0.79%
Colorado	33	0.29%	New Mexico	8	0.28%
Connecticut	15	0.42%	New York	189	-2.43%
Delaware	10	-0.15%	North Carolina	78	-0.25%
District of Columbia	8	-0.14%	North Dakota	1	0.19%
Florida	185	-1.64%	Ohio	73	0.33%
Georgia	69	0.18%	Oklahoma	15	0.54%
Hawaii	10	-0.01%	Oregon	38	-0.4%
Idaho	7	0.23%	Pennsylvania	91	-0.12%
Illinois	105	-0.78%	Rhode Island	6	0.06%
Indiana	42	0.19%	South Carolina	28	0.33%
Iowa	15	0.3%	South Dakota	2	0.18%
Kansas	14	0.27%	Tennessee	36	0.49%
Kentucky	27	0.17%	Texas	175	1.09%
Louisiana	18	0.62%	Utah	7	0.67%
Maine	12	-0.12%	Vermont	4	0.01%
Maryland	54	-0.55%	Virginia	73	-0.63%
Massachusetts	58	-0.47%	Washington	54	-0.07%
Michigan	54	0.65%	West Virginia	6	0.28%
Minnesota	25	0.61%	Wisconsin	43	-0.13%
Mississippi	12	0.38%	Wyoming	3	0.04%
Missouri	35	0.32%			

Table A.4: Age and Gender Representation

Age >=	<	Census Both	Census Male	Census Female	Sample Both	Sample Male	Sample Female
15	19	8.09%	8.39%	7.81%	3.23%	3.41%	2.77%
20	24	8.25%	8.53%	7.98%	21.3%	21.3%	20.5%
25	29	9.03%	9.38%	8.7%	20.1%	19.4%	20.6%
30	34	8.51%	8.7%	8.33%	19.6%	21.6%	17.8%
35	39	8.32%	8.46%	8.19%	13.1%	14.7%	12%
40	44	7.6%	7.66%	7.54%	8.25%	9.05%	7.76%
45	49	7.9%	7.95%	7.84%	5.41%	4.82%	6.1%
50	54	7.9%	7.9%	7.9%	3.79%	2.23%	5.43%
55	59	8.21%	7.99%	8.42%	2.06%	1.29%	2.88%
60	64	7.99%	7.81%	8.16%	1.62%	0.823%	2.44%
65	69	6.74%	6.52%	6.94%	1.06%	1.18%	0.998%
70	74	5.48%	5.32%	5.64%	0.335%	0.118%	0.554%
75	79	3.63%	3.37%	3.88%	0.112%	0.118%	0.111%
80	84	2.35%	2%	2.68%	0%	0%	0%

Using the CCES data (Ansolabehere, Schaffner, and Luks 2020) and entropy balancing (Hainmueller 2012), I also attempted to simulate the results after adjusting covariate balance to a nationally representative sample. Due to the lack of elderly respondents in our sample, some of the resulting weights were over 700, making the results of this model to be meaningless.

Randomization Check and Time-of-Day Effects

The causal identification of the treatment relies on there not being any systematic differences between the treatment (targeted) and control (untargeted) groups. That the control data had to be gathered prior to the treatment step leaves open the possibility of bias due to the difference in time of day. In order to mitigate this bias, the entire experiment was run in as small a window as possible.

The results of Chi-Squared tests of independence on assignment to treatment or control against all of the pre-treatment covariates are presented in Table: A.5. All hypotheses fail to achieve significance except for ideology, but this is not robust to the Holm (1979) or Benjamini-Hochberg (1995) multiple comparisons corrections. Given the strong correlation between partisanship and ideology ($\rho = 0.747$), I find it unlikely that time-of-day effects can confound ideology while not interacting with partisan identification. I therefore conclude that there was a successful randomization, but for each model I additionally test a variant controlling for all pre-treatment covariates.

Table A.5: Chi-Squared Test of Treatment on Pre-Treatment Independence, with Holm and Benjamini-Hochberg corrections.

	p	Holm	BH
Age	0.345066	1	0.7111003
Gender	0.9753692	1	0.9753692
Race	0.5646555	1	0.8873158
Income	0.9483788	1	0.9753692
Region	0.234541	1	0.7111003
NewsInt	0.2219223	1	0.7111003
On-Track?	0.9516303	1	0.9753692
Party	0.3878729	1	0.7111003
Ideology	0.02123161	0.2335477	0.2335477
General Vote	0.7472937	1	0.9753692

A second randomization check took the form of covariate balance checks (see figure A.1). For no variable did the (standardized) mean difference exceed 0.05.

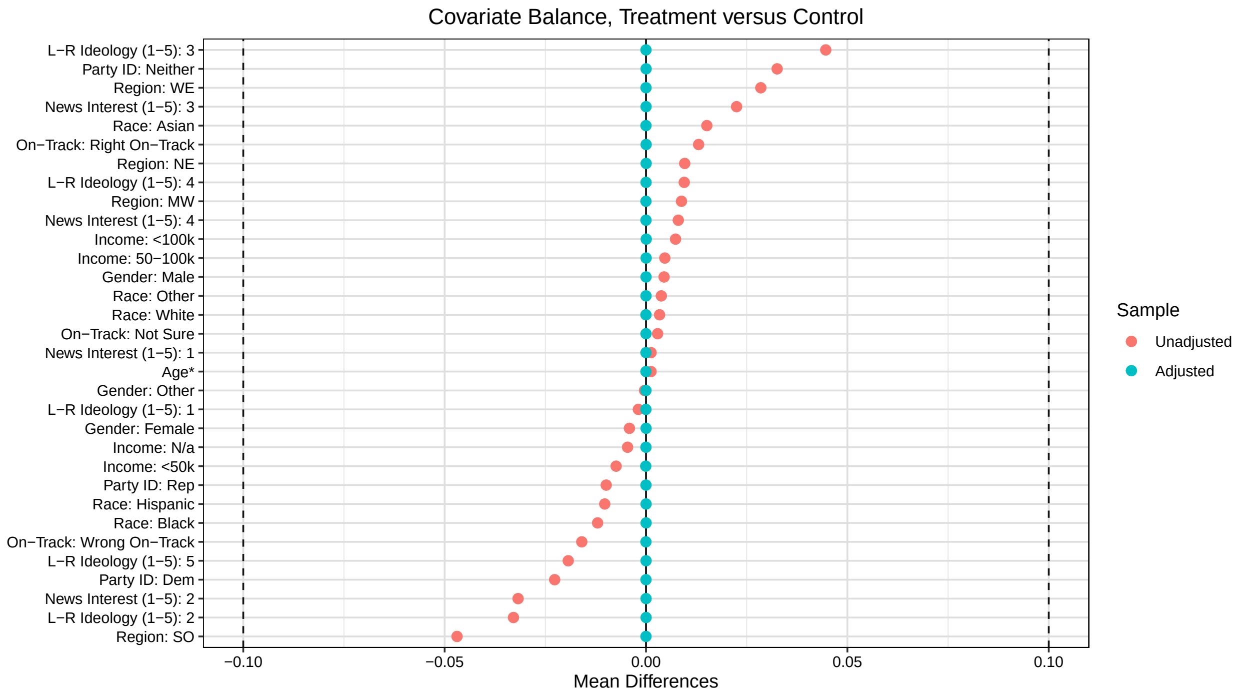

Figure A.1: Covariate Balance on Treatment Indicator

A third randomization check took the form of a logistic regressions predicting the advertisement assignment as a function of pre-treatment covariates. The coefficients are displayed in A.6. It can be seen that no feature is a significant predictor of assignment to treatment or control.

Table A.6: Logistic Regression: Predicting Stage as Function of Pre-Treatment Covariates

Intercept	-0.26 (0.28)
Age	-0.00 (0.00)
Gender: Female	0.00 (0.09)
Gender: Neither	-0.03 (0.32)
Race: Asian	0.08 (0.14)
Race: Black	-0.12 (0.16)
Race: Hispanic	-0.16 (0.18)
Race: Other	0.06 (0.21)
Income: <50k	-0.01 (0.10)
Income: >100k	0.04 (0.15)
Income: N/A	-0.19 (0.29)
Region: Northeast	-0.02 (0.14)
Region: South	-0.17 (0.13)
Region: West	0.07 (0.14)
News: Most Days	-0.11 (0.15)
News: Sometimes	-0.21 (0.14)
News: Hardly	-0.00 (0.24)
Nation: On-Track	0.19 (0.20)
Nation: Wrong Track	-0.02 (0.16)
Party: Dem.	-0.04 (0.11)
Party: Rep.	-0.10 (0.19)
Ideo: Lib.	-0.11 (0.12)
Ideo: Neither	0.18 (0.16)
Ideo: Con.	0.03 (0.18)
Ideo: Very Con.	-0.45 (0.27)
AIC	3013.90
BIC	3156.99
Log Likelihood	-1481.95
Deviance	2963.90
Num. obs.	2261

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Appendix B: Theoretical Notes

Predictive Model

As mentioned in the main article, our claims of successful emulation of micro-targeting depend in part on the targeting model being sufficiently accurate. There are two scenarios where the treatment effect would not be attributable to optimized allocation. First is zero heterogeneity; if the relative effectiveness of the advertisements is constant across respondents, then the change in average favorability may be due to allocating a greater proportion of respondents a more effective advertisement. This is addressed in the main article, as well as a subsequent section of this appendix. The second is poor predictions, in which case the model cannot be credited with the improvement. Neither of these situations could convincingly be described as emulating microtargeting.

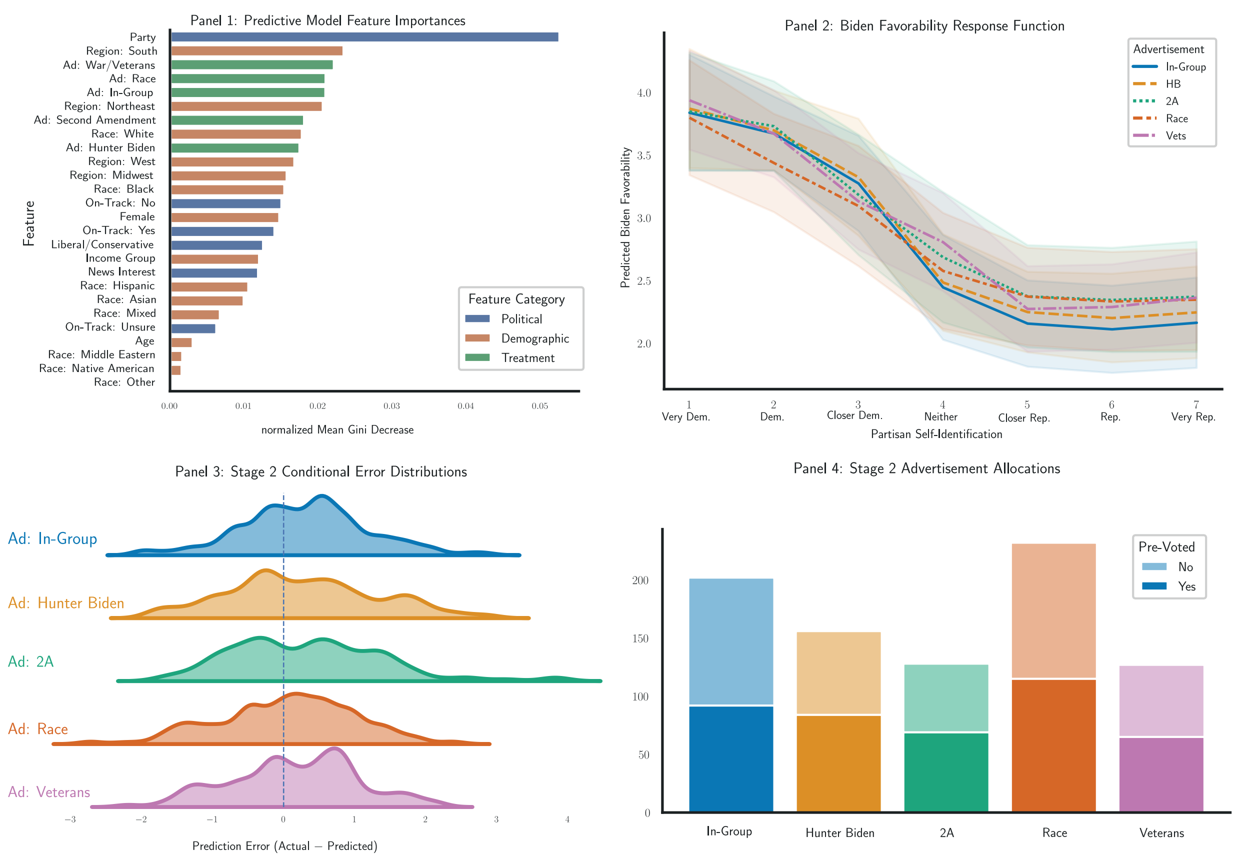

Figure A.2: *panel 1: feature importances for predictive model used to allocate optimal advertisements; panel 2: predicted Biden favorability response curve as function of partisanship for stage 2 respondents; panel 3: conditional prediction error distributions; panel 4: stage 2 advertisement allocation counts.*

Figure A.2 addresses these objections. Panel 1 reports per-feature normalized Mean Gini Decrease[3] (nMGD). Note all advertisements have high nMGD, indicating that the algorithm "learned" more about a respondent's Biden favorability from the advertisement they were shown than their gender, income group, or race (with the exception of *Race: White* versus *Ad: Hunter Biden*). Although feature importance does not correspond to any causal estimand, the randomized advertisement assignment in this first stage leaves no explanation for high feature importance other than the heterogeneous effects between advertisements.[4]

[3]See technical explainer (SI-D).

[4]A causal interpretation of stage 1 results is discussed below.

The favorability-partisanship response curves in panel 2 provide further insight into the logic of the targeting model. This panel shows predicted Biden favorability for each advertisement as partisanship is varied, averaged across all stage 2 respondents.[5] The overlapping curves demonstrate between-advertisement and between-respondent heterogeneity over partisanship; one advertisement (*Race*) is more effective on average when the voter is Democrat, while a different advertisement (*In-Group*) is more effective on average when Republican. The large overlapping standard errors indicate considerable level differences on between-respondent heterogeneity for variables other than partisanship and advertisement.

The conditional error distributions on stage 2 predictions (panel 3) exhibit a weak positive bias ($\mathbb{E}(e) = 0.21$, $p < 0.001$), indicating that on average the model over-estimated the effectiveness of the advertisements. Interpreting this value is difficult because the realization of predicted outcomes is confounded by the model itself. Note the *2A* advertisement has a long right tail, reflecting its greater assignment to individuals with a lower baseline level of Biden favorability, increasing the magnitude of possible prediction error.

It is unclear whether bias is a cause for concern; what is important is that the algorithm did not selectively favor one advertisement. The first three advertisements demonstrate similar conditional average prediction errors (0.31, 0.28 and 0.34 respectively), reducing the likelihood of distorting the ranking between each other. Crucially, panel 4 demonstrates that no single advertisement dominated the allocations, reducing the likelihood that the difference between treatment and control is due to a more effective advertisement being shown to more respondents.

Characterization

While panel 1 of figure A.2 tells us which features mattered, it does not tell us how they mattered. In order to characterize the kinds of respondents that were shown each advertisement, I fit a linear regression for each advertisement to predict assignment to that advertisement as a function of pre-treatment traits.[6]

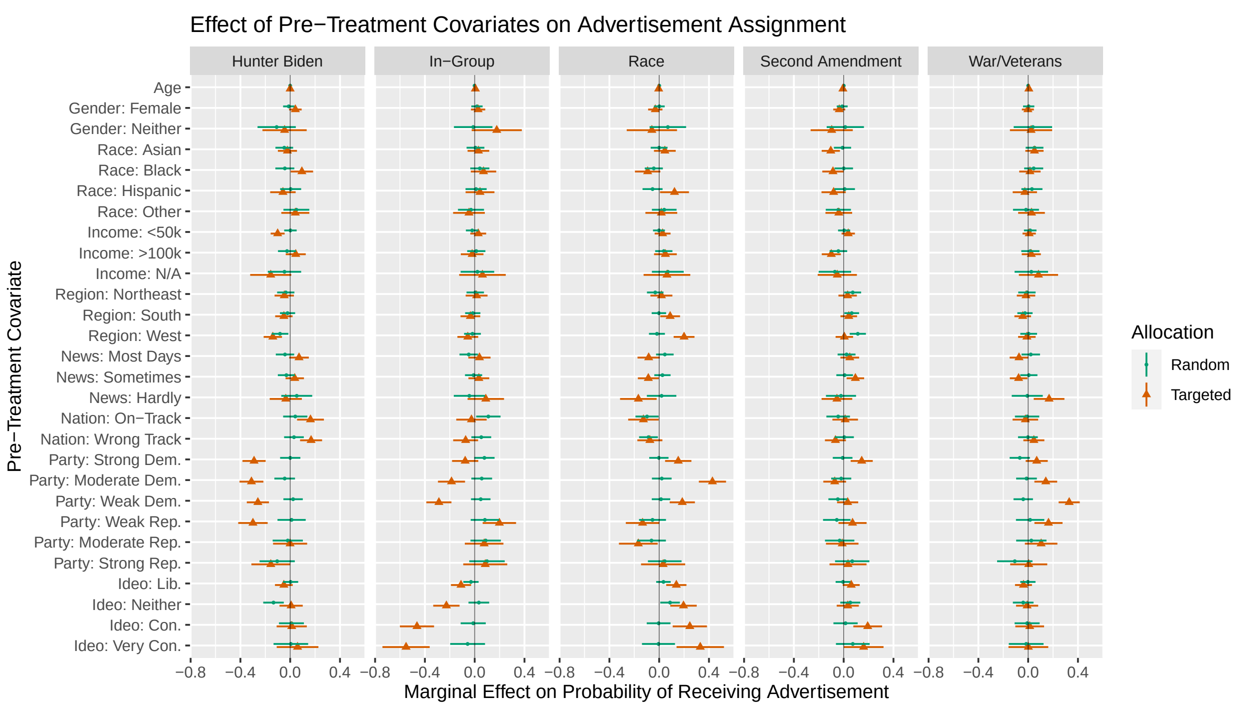

Figure A.3: coefficient plot of ten linear regressions predicting advertisement assignment

[5]See SI-D for interpretation.

[6]In simpler terms: to fit the model predicting the effect of pre-treatment traits on receiving the Hunter Biden advertisement, a new dummy variable was created that took the value 1 if the respondent received the Hunter Biden advertisement, and 0 otherwise.

Each panel of the figure represents an advertisement. The red horizontal bars with a triangular marker at the center show the coefficient with 95 percent confidence intervals for the targeted sample, and the blue horizontal bars with the circular marker show the coefficient and 95 percent confidence intervals for the sample receiving randomized advertisements. These coefficients can be interpreted as the average effect of a one-point change in the corresponding pre-treatment variable denoted on the y-axis label on the probability of receiving the advertisement in the label at the top of the panel.

For instance, looking at the red bars, we can see that relative to being completely unaligned, identifying as a Democrat or a weak Republican reduced the probability of receiving the Hunter Biden advertisement by roughly 0.3. A few patterns are unintuitive. In particular, it appears that respondents identifying as ideologically conservative were more likely to receive the Race advertisement.

The blue bars provide evidence of sample balance in the first stage; given that advertisement allocation was randomized, it should not be predictable from any pre-treatment covariates.

Decomposition

As noted in the design and results section, the average treatment effect of targeting in this experiment can be decomposed into two parts; the change due to between-advertisement heterogeneity, and the change due to within-individual heterogeneity. In this section I explain this decomposition and show how the results translate.

In theory, it is possible to control for this effect by fixing the proportion of advertisements between samples. I opt not to do this because a) this is not a realistic constraint that campaigns face, b) the online bipartite matching with stochastic vertex arrivals problem still has no globally optimal solution (see Goyal and Udwani 2020 for current state of the literature) and c) these issues can be addressed post-hoc with the following simple decomposition.

Denoting the average treatment effect as τ, I note that it can first be decomposed into the sum of per-advertisement outcomes times the probability of receiving that advertisement:

$$\tau = \sum_{j \in A} (p_{j2}\bar{y}_{j2}) - (p_{j1}\bar{y}_{j1})$$

Here $A = [a, b, c, d, e]$, the set of the five advertisements, and the second subscript $[1, 2]$ indicates the stage of the experiment. Note that 2 is the treatment stage.

Rearranging the terms, we get

$$\tau = \sum_{j \in A} (p_{j2} - p_{j1})\bar{y}_{j1} + (\bar{y}_{j2} - \bar{y}_{j1})p_{j1} + (p_{j2} - p_{j1})(\bar{y}_{j2} - \bar{y}_{j1})$$

This can be broken down into three parts:

- $(p_{j2} - p_{j1})\bar{y}_{j1}$: the change in advertisement proportions weighted by the stage 1 outcomes.
- $(\bar{y}_{j2} - \bar{y}_{j1})p_{j1}$: the change in outcome weighted by the (uniform) stage 1 frequencies.
- $(p_{j2} - p_{j1})(\bar{y}_{j2} - \bar{y}_{j1})$: an interaction term.

The first term represents the change due to a greater proportion of individuals being shown a given advertisement; in other words, the change due to between-advertisement heterogeneity. The second term represents the change in individual outcomes for each advertisement between stages; in other words, the effect of allocating individually better advertisements.

This table shows that 98 percent of the increase in Biden dislike was attributable to changes in individual outcomes; in other words, the effect of showing respondents individually best advertisements. Likewise, roughly 2 percent of this change is attributable to changing the mixture of advertisements. This pattern holds across models, where in all cases the change is almost entirely attributable to within-individual heteroegeneity.

Table A.7: Effect Decomposition for Four Models

	Dislike Biden Total Sample	Unaligned and Non Pre-Voting	Vote Biden Total Sample	Unaligned and Non Pre-Voting
$(p_2 - p_1)y_1$	0.006649	0.006460	0.006649	-0.001744
$(y_2 - y_1)p_1$	0.034626	0.085389	0.034626	-0.069409
$(y_2 - y_1)(p_2 - p_1)$	-0.011849	-0.005128	-0.011849	0.000316

Relation to Coppock, Hill, and Vavreck (2020)

This paper makes extensive reference to Coppock, Hill, and Vavreck (2020) (within this section of the appendix, CHV20). CHV20 and its replication materials, published just over a month prior to this experiment, were essential to its design and testing. The replication data was used as mock results for the first stage, which were then used to train predictive algorithms and simulate the targeting in stage two.

The simulated targeting experiment based on the CHV20 data indicated that targeting would have an effect, with a magnitude of roughly 0.4 on a 1-5 scale. This result, however, is difficult to interpret because the expected outcome under targeting was given by the same model that was used to optimize advertisement allocations. To increase certainty that this result was not an artifact of the algorithm selectively sampling off the right-hand-side of the standard error, I conducted a permutation test ($n = 30,000$) in which the treatment vector was randomized. The null hypothesis being tested by this permutation test was row-level treatment independence, which would make the targeting irrelevant. This null hypothesis was rejected with $p = 0.99$.

We subsequently used the estimated effect size and variance as the basis for a pre-experimental power analysis, shown in Figure A.4. On this basis I removed two additional treatment categories and a control advertisement, which I intend to test in a follow-up experiment.

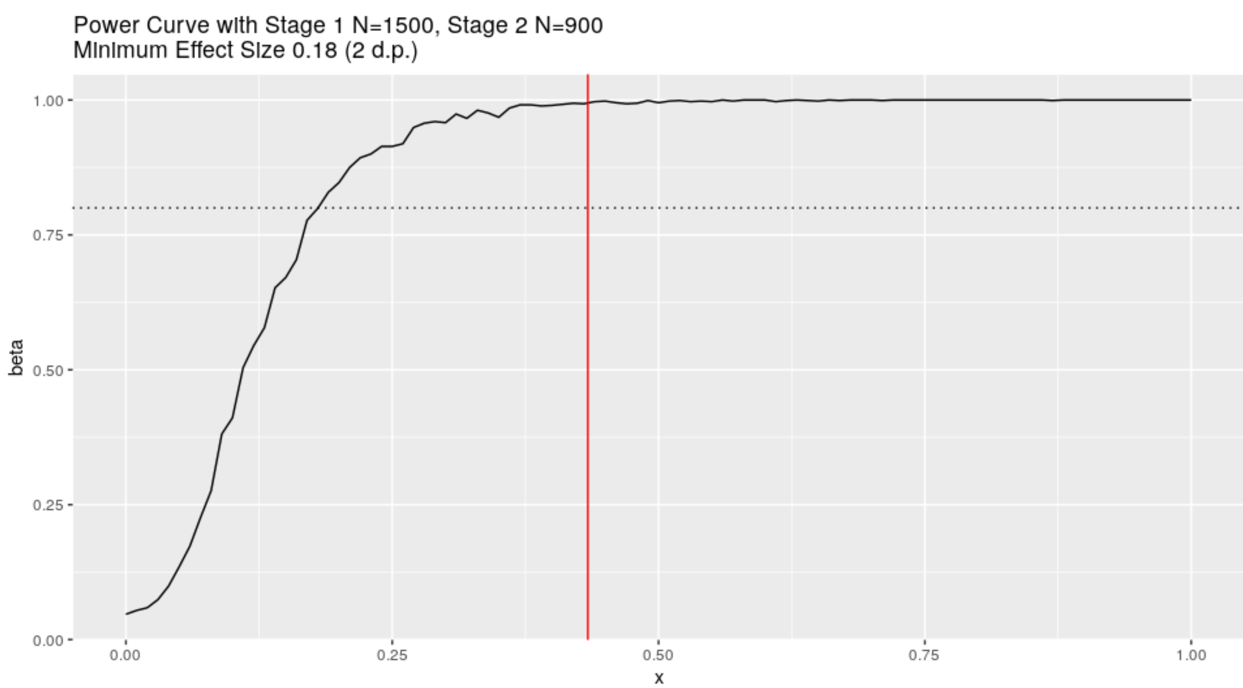

Figure A.4: *power analysis from CHV20 data*

As noted, the conclusions of this paper seem at odds with CHV20, who find little evidence of treatment effect heterogeneity. They infer from this that there is little room for micro-targeting to work; if the effect of advertisements does not vary greatly from one individual to another, why should micro-targeting make any difference?

We suggest that there are two main reasons for this difference. The first is that whereas CHV20 focuses primarily on heterogeneity over advertisements, this experiment maximizes heterogeneity within voters. Thus, although CHV20 search for conditional effects over the characteristics effect of 49 different advertisements, the only respondent characteristic they test for heterogeneity over is partisanship.

In comparison, the targeting algorithm trained for this experiment searches a space of up to $362,880$ (9!) combinations of respondent characteristics to find connections between response to the advertisement and combinations of traits. The respondent is then shown the advertisement predicted to maximize (or minimize) this difference. As referenced in the main body, CHV20 does indeed note that exhaustively searching traits may expose sources of heterogeneity.

Advertisement Effectiveness

A natural estimand of interest is the “effect” of each advertisement. In contrast to other papers in this literature, this paper does not offer an estimate of the persuasive effect of each advertisement.

A theoretical quantity of interest is the effect of each individual advertisement used in this experiment. Randomized advertisement assignment in the first stage allows for a causally identified interpretation of the coefficient on each advertisement, but because it is unclear which advertisement should serve as a reference category, these coefficients are not meaningful. The original design for the experiment included a control advertisement in stage 1 in order facilitate these comparisons, but this was omitted to prioritize the key components of this experiment: maximizing the number of observations for training the predictive algorithm and identifying the effect of targeting. The implementation of the control group can be seen in the source code of this project on GitHub.

Envelope Calculation

The reader may be wondering what are the substantive implications of the estimated effect size. The experiment shows that in a hypothetical campaign where the Trump campaign ran just five advertisements, optimizing allocation with the method presented in this paper would have resulted in 8.7 percentage points more of unaligned voters stating that they do not like Biden, and 7.1 percentage points fewer of unaligned voters stating that they intend to vote for Biden, when compared to a scenario when advertisements are assigned independently of individual traits.

What do 8.7 and 7.1 percentage points mean? While it is not possible with the available data to give a rigorous estimate, the following calculation gives an illustration of how meaningful a 7.1 percentage point swing could be. In the 2020 United States presidential election, 31 states included information on partisan affiliation for voter registration. Assuming (unrealistically) that all registered voters turn out to vote, that stated intention not to vote for Biden translates into voting for neither candidate (i.e. no switching), then we can multiply the proportion of unaligned voters in each state by the effect size in order to estimate how the result would have changed.

For example, in Florida, where 26.7% of registered voters affiliated with neither of the major parties, a swing of 7.1 percentage points among unaligned voters could lead to a diminution of $0.267 \times 0.0708 = 0.0189$, or 1.9 percentage points in the Biden vote share. In North Carolina and Arizona, this estimated change is larger than the percentage point difference between the shares of the two leading candidates (in Arizona, 35.1% unaligned with a 0.3% margin of victory, and in North Carolina 30.6% of voters are unaligned with a 1.3% margin of victory).

There are many obvious caveats and limitations to this calculation beyond the unrealistic assumptions already stated. For one, changing answers on a survey is radically different to changing voting intention. For another, this survey design and calculation do not account for decay and counter-acting effects, meaning the actual effect is likely smaller. A further point is that the 7.1 percentage point effect is based on a distribution of Biden preferences that were present in the sample; this is likely not the same for all possible distributions of starting preferences. Finally, the sample used in this survey is not representative, meaning that the distribution of effects will likely be different. Nevertheless, this calculation is meant to illustrate that the magnitude of effect is not trivial in a political context like the United States, where consequential races have incredibly small margins.

Normative Implications

These results highlight the potential for a multitude of harms. Arguably the clearest threat is to voter privacy, which this technology creates incentives to disregard (Wachter 2017). Moreover, because those using this technology are more likely to win power, this creates a self-reinforcing cycle (Hindman 2018, 5) in which private actors (Zuboff 2015; Baldwin-Philippi 2017) and public officials (Krotoszynski 2020) broker access to such data.

There is a second, less clear, set of normative issues that speaks to manipulation and the legitimacy of democratic processes. It begins with asking "what does it mean to show a voter the *right* advertisement?" Proponents of targeted advertising often present it in a benign and efficiency-improving frame; targeted advertising reduces the informational burden by showing the consumer only the goods that are most relevant to their interests (Hindman 2018, 39). Analogously, targeted political advertising has the potential for good; it can be used to show voters elements of candidates' policies that are most relevant to their interests, lowering the cost of civic and political engagement.

Targeted advertising can also be malevolent. Combining emotive advertisements with knowledge of the fears and anxieties of the voters[7], targeted negative advertising could fairly be called a form of manipulation (Sunstein 2015). By seeking to identify ways to elicit strong negative reactions from voters, campaigns are attempting to engage peripheral, or emotive, and not central, or cognitive/logical, processing of the information in advertisement. This approach to advertising attempts to bypass the deliberative and cognitive capacity of the viewer, ultimately seeking to make the decision for the viewer.

This negative and manipulative strategy is a more suitable description of the calculus employed by CA and demonstrated in the five Trump campaign advertisements used in this experiment. The psychometric profiling employed by CA targeted "neurotic" voters that would be especially susceptible to fear-based advertising (Hindman 2018). The five advertisements similarly appear to be attempting to elicit negative emotions to be associated with Biden, from the fear that he will take away the viewer's guns, to emphasizing that Biden and Clinton laugh at and mock people like the viewer. To determine the extent to which this strategy characterizes the various United States campaigns requires further research.

However, these are issues with manipulation and not targeting. At an individual level, manipulative political advertising could be seen as disenfranchising, by depriving voters of the opportunity to make their own decision. The results from this experiment show that such manipulation is possible on a massive scale, and can therefore affect the outcome of an election. That election outcomes could be determined by micro-targeted manipulative advertisement campaigns undermines the legitimacy derived from the democratic process, as such outcomes arguably no longer reflect the public will. In a nutshell, although campaigns can engage in manipulative practices without targeting, to the extent that a manipulative campaign is effective at a societal level the threat transcends from individual to systemic.

[7]Note that with the black box algorithmic approach to targeting employed here, the advertisers may not specifically know what aspects of an advertisement make it resonate with an individual. This does not mean, however, that the algorithm is not implicitly leveraging these patterns or identifying these psychological traits.

Appendix C: Regression Tables and Robustness

The following appendix contains the full results of the various models and operationalizations, as well as various robustness checks.

All Operationalizations and Specifications

In this section of the appendix, the various regression models reflect different methods of operationalizing the outcome. As these decisions should not, and were not made arbitrarily, we explain the details and rationale here. For the candidate favorability question, after watching the advertisement, survey respondents were posed the following question:

> *"On a scale of 1-5, how would you rate Democratic Presidential Candidate Joseph Biden? (A rating of 4 or 5 means that you feel favorable and warm towards the candidate. A rating of 1 or 2 means that you don't feel favorable and warm towards the candidate. You would rate them at 3 if you don't feel particularly warm or cold toward them.)"*

This item is operationalized in four ways. The first two keep the five categories, but treat it either as a continuous linear scale, or as an ordinal categorical scale. The latter two treat it as a binary indicator of Biden dislike, with 1-2 being coded as 0, and 3-5 being coded as 1. The direction of the measure and inclusion of 3 in the positive category are not arbitrary; Biden *dislike* should be of particular interest to campaigns. If they can get voters to actively dislike the candidate (1-2), then this is more significant than simple indifference (3).

The vote choice question simply allowed them to indicate whether they intended to, or had already voted for Trump, Biden, another candidate, that they did not intend to vote, or that they did not wish to answer. The intent to vote Biden variable is simply a binary indicator on whether they chose "I intend to vote for Biden" for this question.

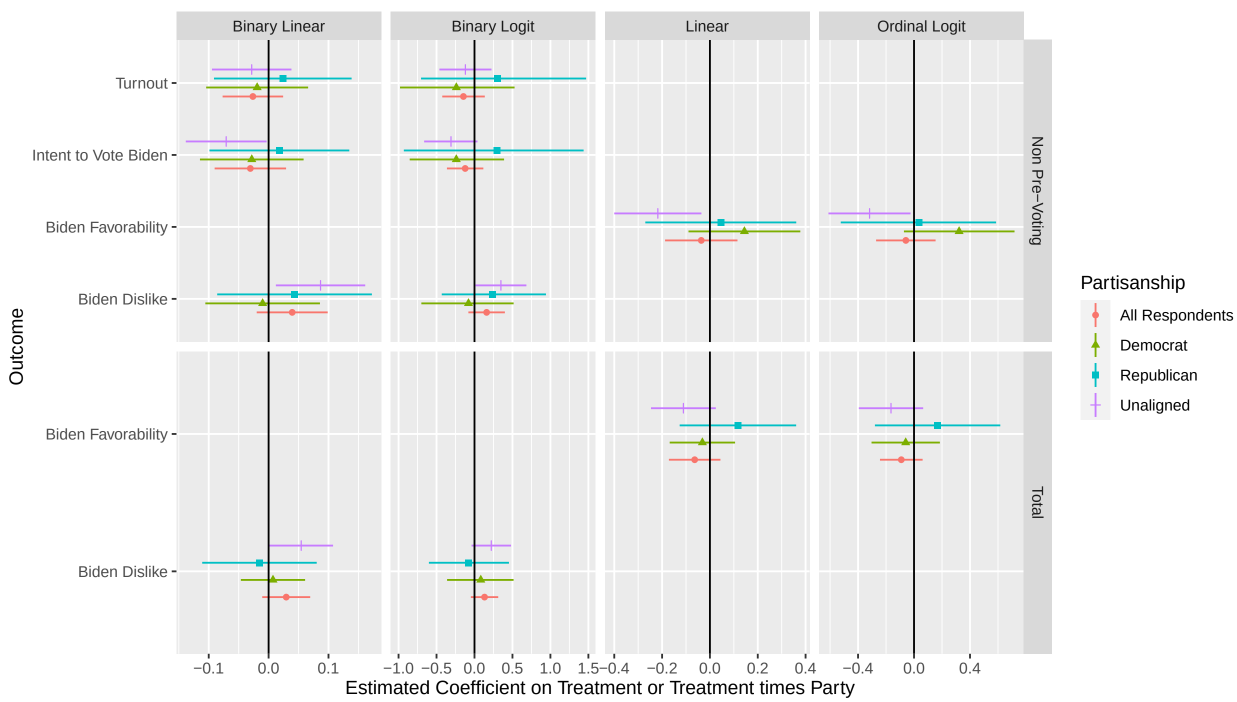

Figure A.5: *coefficient and 95% confidence interval on treatment indicator for all specifications*

Each dot-and-whisker in figure A.5 shows the coefficient and 95% confidence interval on the treatment indicator for one of 48 different models. The top four panels contain models fitted to the subset of respondents who had not voted prior to the survey ($N = 1,160$), whereas the bottom four panels contain coefficients for the full sample ($N = 2,261$). The left-to-right, the panels show linear models fit to a binary outcome (Binary Linear), logit models fit to a binary outcome (Binary Logit), linear models fit to a continuous outcome, and an ordinal logit model fit to a Likert outcome. The ticks on the y-axis designate the outcome of the model; *Turnout* is a binary indicator on whether the respondent stated that they intend to cast a vote in the election, *Intent to Vote Biden* is a binary indicator on whether the respondent stated that they intended to vote for Biden, *Biden Dislike* is a binary indicator rating Biden lower than "Neither" higher on a favorability scale, and *Biden Favorability* Likert scale ranging from "Strongly Dislike" to "Strongly Like."

The coefficients are presented in groups of four, with each shape/color corresponding to which covariates were included in the model. The red circle indicates the coefficient on treatment for a model regressing only treatment on outcome. The green triangle indicates the coefficient on treatment for a model regressing treatment interacted on partisanship with Democrats set to the base partisanship category. The blue square and purple vertical line indicate the same with Republicans and Unaligned voters set as the base partisanship category respectively.

Effect on Candidate Favorability

Table A.8 reports the effect of targeting conditional on partisanship for the subset of respondents who had not already voted at the time of the survey ($N = 1,160$). The coefficients in this table reveal a substantial amount of treatment effect heterogeneity. The second row of coefficients, *Targeting (Unaligned)*, shows the effect of targeting on respondents who had not already cast their votes and self-identify as neither party; the group that a campaign would aim to target. The estimated CATE is -0.218 points on a five-point scale ($SE = 0.093$), *which translates to an* 8.7 *percentage point increase in respondents saying they dislike Biden* ($SE = 3.8$pp). Both of these coefficients are significantly different from zero at standard confidence levels of $\alpha = 0.05$ ($p = 0.0192,\ 0.0338,\ 0.0228,\ 0.0416$).

Table A.8: Effect of Micro-Targeting on Candidate Favorability, Interacted on Partisan Self-Identification among Respondents who had not Voted

	OLS: Five-Point	Ordered Logistic	OLS: Binary	Logistic
Intercept	2.588^{***}		0.467^{***}	-0.133
	(0.057)		(0.024)	(0.105)
Targeted (Unaligned)	-0.218^{**}	-0.318^{**}	0.087^{**}	0.348^{**}
	(0.093)	(0.150)	(0.038)	(0.171)
Democrat	0.997^{***}	1.625^{***}	-0.314^{***}	-1.580^{***}
	(0.092)	(0.159)	(0.038)	(0.212)
Republican	-0.693^{***}	-1.276^{***}	0.269^{***}	1.159^{***}
	(0.108)	(0.189)	(0.044)	(0.216)
Targeted × Democrat	0.362^{**}	0.640^{**}	-0.097	-0.427
	(0.151)	(0.251)	(0.062)	(0.353)
Targeted × Republican	0.263	0.352	-0.043	-0.112
	(0.186)	(0.320)	(0.076)	(0.388)
R^2	0.258		0.190	
Adj. R^2	0.254		0.187	
Num. obs.	1160	1160	1160	1160
AIC		3259.343		1363.093
BIC		3304.849		1393.430
Log Likelihood		-1620.671		-675.547
Deviance		3241.343		1351.093

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

The remaining coefficients reveal unsurprising patterns. The effect of self-identifying as Democrat and Republican has large and significant effects on Biden favorability in the expected directions. Looking at the CATEs of self-identification as Democrat or Republican, we discover that their coefficients are in the same direction and diminish the effect of targeting. In other words, targeting anti-Biden advertisements has the strongest effect on unaligned voters, but has a relatively weak effect on voters who identify with a party.

Effect on Voting Preference

The dependent variable for the models in this section is the proportion of respondents stating their intention to vote for Biden in the general election, out of the respondents who had not voted at the time of the survey. Table A.9 reports the effect of targeting on intention to vote for Biden among respondents who had not voted at the time of the survey. The two columns on the left report the models regressing targeting on voting preference, and the two columns on the right report the models regressing targeting interacted on partisan self-identification on voting preference. The columns alternatingly report the results for OLS and logistic regression models.

Table A.9: Effect of Micro-Targeting on Intention to Vote for Biden among Respondents who had not Voted

	OLS Uninteracted	Logit Uninteracted	OLS Interacted	Logit Interacted
Intercept	0.478^{***}	-0.090	0.392^{***}	-0.438^{***}
	(0.018)	(0.074)	(0.021)	(0.108)
Targeted (Unaligned)	-0.030	-0.123	-0.071^{**}	-0.309^{*}
	(0.030)	(0.122)	(0.034)	(0.179)
Democrat			0.485^{***}	2.409^{***}
			(0.034)	(0.229)
Republican			-0.337^{***}	-2.395^{***}
			(0.040)	(0.379)
Targeted × Democrat			0.043	0.070
			(0.056)	(0.363)
Targeted × Republican			0.089	0.609
			(0.069)	(0.617)
R^2	0.001		0.346	
Adj. R^2	0.000		0.343	
Num. obs.	1160	1160	1160	1160
AIC		1605.846		1158.461
BIC		1615.958		1188.798
Log Likelihood		-800.923		-573.230
Deviance		1601.846		1146.461

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

The first pair of models (1) and (2) show a weak, insignificant and negative ATE of targeting on voting preference. When we search for heterogeneity over (non-) partisanship, we observe a similar pattern to the CATE of targeting on candidate favorability. The effect of targeting on intention to vote for Biden is -0.071 ($SE = 0.034$, $p = 0.03967$), meaning that among respondents who identified with neither party and had not voted at the time of the survey, being targeted increased the proportion of respondents not intending to vote for Biden by 7.1 percentage points. As with the previous section, the effect of aligning with either the Democratic or Republican party largely nullifies the effect of targeting. That the pattern is persisting in a separate but related outcome increases confidence that targeting is in fact increasing the likelihood that the targeted advertisements are persuasive.

Effect on Turnout

The final set of results, shown in table A.10, looks at the effect of targeting on turnout on respondents who had not voted yet. This is operationalized as a dummy variable indicating the proportion of respondents stating that they will vote for Biden, Trump, or a third candidate. The models are presented in the same as the previous section, with the first two columns testing an uninteracted model and the latter two testing a model interacting turnout intention on partisan self-identification.

Table A.10: Effect of Micro-Targeting on Turnout among Respondents who had not Voted

	OLS Uninteracted	Logit Uninteracted	OLS Interacted	Logit Interacted
Intercept	0.777^{***}	1.250^{***}	0.628^{***}	0.523^{***}
	(0.015)	(0.086)	(0.020)	(0.105)
Targeted (Unaligned)	-0.021	-0.118	-0.019	-0.082
	(0.025)	(0.139)	(0.033)	(0.170)
Democrat			0.298^{***}	2.002^{***}
			(0.032)	(0.267)
Republican			0.285^{***}	1.821^{***}
			(0.037)	(0.299)
Targeted × Democrat			0.004	-0.118
			(0.053)	(0.416)
Targeted × Republican			0.035	0.303
			(0.065)	(0.568)
R^2	0.001		0.126	
Adj. R^2	-0.000		0.123	
Num. obs.	1241	1241	1241	1241
AIC		1343.142		1184.504
BIC		1353.389		1215.246
Log Likelihood		-669.571		-586.252
Deviance		1339.142		1172.504

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

These models indicate that if there is an effect of targeting on turnout, then it is not significantly different from zero in the total sample nor among non-partisan voters. It is worth noting that partisan voters were more likely to indicate that they intended to vote by 28.6 and 26.7 percentage points for Democrats and Republicans respectively, from a baseline of 63.5% for non-partisan voters.

Appendix D: Technical Explainers

This appendix contains brief explanations of technical concepts not assumed to be known by a reader with a political science background.

(normalized) Mean Gini Decrease

Unlike ordinary least squares (OLS), the random forest model employed in the second stage does not have regression coefficients to facilitate interpretation of the predicted outcome as the sum of linear components. Nevertheless, decision-tree based algorithms do provide metrics that allow substantive interpretation (Montgomery and Olivella 2018).

Panel 1 of figure A.2 article shows the per-feature normalized Mean Gini Decrease (nMGD) statistic from the RF model used to assign allocations in the second stage. nMGD tells us the degree to which partitions on a feature produce homogeneous sub-spaces at each node of the individual decision trees. In other words, this statistic tells us the extent to which a feature produces systematic and separable regions of the label space. Because Mean Gini Decrease is biased upwards for features with high cardinality, we normalize the score by dividing by cardinality.

A question of substantive interest to both researchers and campaigners alike is why the predictive algorithm assigned a particular advertisement and not another. Features with high nMGD provide much of the predictive power for RF models. Partisan self-identification has the highest feature importance, followed by the respondent being from the South.

Average Response Curve

Panel 2 of figure A.2 presents, for each advertisement, the average response curve of the outcome (Biden favorability) across values of partisanship on a seven-point scale. This was calculated by using the fitted RF model to predict Biden favorability for each of the stage 2 respondents at each level of partisanship (1 through 7), and then averaging the favorability across respondents for each partisanship-advertisement combination. The shaded areas represent the standard deviation of the predicted response curves for each advertisement.

Focusing just on the solid lines (and not the shaded areas), the interpretation of the figure is as follows. At each level of partisanship, the vertical ordering of the lines represents the relative average predicted effectiveness of the various advertisements. For example, at partisanship *2: Dem*, the line corresponding to the *Race* advertisement is lower than the other four. This indicates that on average, for the stage 2 respondents, if we set partisanship to *2* on the seven-point scale and leave all other traits as they were, the targeting algorithm predicts that showing the *Race* advertisement will produce a lower Biden favorability other advertisements. Similarly, when we move up the scale to partisanship of *5*, *6* or *7*, the *In-Group* advertisement is predicted to produce a lowest Biden favorability on average, followed by the *Hunter Biden* advertisement.

The shaded areas, as mentioned, indicate the standard deviation of each of these response curves. You may note that they are largely overlapping, perhaps with the exception of *Race* at partisanship *2*. The takeaway here is that there is still a great deal of heterogeneity in individual response curves (the predicted levels of Biden favorability for each individual, artificially varying their partisanship). This should not be surprising; we would expect other demographic traits, such as state of residence, to correspond with level shifts in baseline levels of Biden favorability. Note also that artificially varying partisanship results in some very implausible trait combinations (e.g. a young ideologically left-wing hardcore Republican).

Ultimately, the takeaways from this figure are as follows:

- *Crossing Curves*: the algorithm predicts that different advertisements are the most effective at different levels of partisanship. This provides evidence supporting the hypothesis that the effect of micro-targeting is a “substitution” effect, and not an “income” one.
- *Large Overlapping Errors*: the algorithm predicts that there is a high level of heterogeneity in response corresponding with traits other than partisanship. In other words, while partisanship and the advertisement shown matter, they do not explain the full range outcomes.

- *Small Difference at Very Democrat*: that the *Race* advertisement is comparatively much more effective at moderate Democrat partisanship (*2*) than extreme Democrat partisanship (*1*) indicates that there may be no persuasion effect for any advertisement when the respondent identifies as strongly Democrat. This is consistent with the findings of Broockman and Kalla (2020).

Appendix E: Ethics

As an experiment involving mild deception, and dealing with difficult ethical issues, every step was taken to ensure that this research was conducted in a way that respected and protected participants and avoided impacting wider processes such as the ongoing election. Where possible, the researcher went beyond the requirements of the ethical review board and erred heavily on the side of caution.

IRB Approval

Prior to the experiment, ethical approval was applied for and obtained from the REDACTED University Research Ethics Committee, REDACTED Departmental Research Ethics Committee (DREC), with Research Ethics Approval Reference Number `REDACTED`.

Recruitment and Payment

Participants were recruited via academically-focused online survey provider Prolific, who require that researchers provide fair compensation for task completion. Accordingly, respondents were compensated at an average rate of USD 11.51 per hour (with a median completion time of three minutes). A call for responses was posted via the Prolific portal with the following text:

```
Political Ads Survey

Summary:
    - 2-4 minute academic survey.
    - Involves 13 multiple-choice questions and a 30-second ad.
Compatibility:
IMPORTANT: Please disable your ad blocker,
           as some participants have had issues watching the video while it is enabled.
    - Tested on Linux/Mac/Windows for Firefox/Chrome.
    - There are no cookies/external ads on this survey website.
Questions:
    - 5 demographic
    - 4 political
    - 4 about the ad
```

Respondents were therefore given a great deal of up-front information regarding the content and objective of the task, so that they could decide whether they wanted to participate.

Respondents who selected this study were provided with a link to an external website, `survey.<REDACTED>.org` along with a `GET` url parameter to forward their prolific user id. This website was built and hosted by the researcher, and all connections to it were forced to use encrypted `https` in order to minimize the risk of a data leak. The aforementioned url parameter is only linkable to real identities by the survey provider, Prolific.

Consent

The first page of the survey, `https://survey.<REDACTED>.org/consent.php`, explained in broad terms the purpose of the research (researching the effects of political advertising), obtained participant consent, and made clear that respondents had the option of freely revoking their consent at any point with no penalty. It also made clear to participants that a longer debrief would be available at the end of the survey, and that this would contain additional information regarding the purpose of the experiment.

Deception

This design involves deception by omission. The real objective of this study was to study the effect of targeted advertising. Participants did not consent to having their data used to train a targeting algorithm, or being targeted by said algorithm, prior to these things being done.

To address this, at the end of the survey, participants were shown an extensive debrief explaining the intention of the experiment (measuring the effect of targeted advertising) and that either a) their responses had been used to train a targeting algorithm or b) they had been targeted with the advertisement they were shown based on the answers they gave. That participants were able to revoke consent at any point was emphasized again, and the contact details of the researcher were provided again to answer questions and concerns.

Impact

The debrief emphasized that these advertisements were run by the Trump campaign, and that the purpose of this experiment was to show the effect of targeting the advertisement. Research into political microtargeting finds that informing participants of the targeting alters engagement with the message (Kruikemeier, Sezgin, and Boerman 2016), such that revealing the intention of the research is likely to counteract the effect of the advertisement itself.

Moreover, from follow-up discussions with participants, more than 100 indicated that they had been "inundated" with political advertisements in the period leading up to the survey, and as such were unable to remember the content of the advertisement they were shown even half an hour after the experiment. Specifically, many were unable to even recall whether we had shown them an advertisement run by the Trump or Biden campaign.

Confidentiality and Data Privacy

All connections were made to the survey website via secured connection, and all answers were stored locally. Pseudo-anonymised information taken during the course of the survey was a unique user id provided by Prolific, which the researcher does not have the means to convert to real identities.

Responses were securely transferred to the researcher's local device via SQL over SSH. A copy and backup are held by the researcher in password-protected databases. The reproduction data will be scrubbed of any metadata provided by Prolific, to only contain the answers to the questions.

A copy of the consent responses, linked to Prolific IDs, will be held for the required period of time in order to permit the researcher to take action in the case of revoked consent.

Funding

This study was made possible from the researcher's own expenses and a generous grant from the <REDACTED FOR ANONYMITY> fund.

Conflicts of Interest

The author hereby declares that they have no conflicts of interest, personal or professional, relating to this study.

Appendix F: Reproducibility

Reproduction code for both the experimental design (website front end and back end) as well as scripts for analysis will be hosted on Github at `https://github.com/<REDACTED>`. Data will be made available via Dataverse once appropriately sanitized and the period stated in the consent form for revoking consent has expired.

Technical Implementation

The design of this survey required a high degree of interactivity. In the second stage, the respondents' answers were sent to an on-line pre-trained machine learning model that would respond with the optimal advertisement assignment in real time. Given that there was no commercial survey software that provided this integration functionality, the survey website was built by the researcher from the ground up.

The front-end of the website[8] was largely written in PHP and hosted on an AWS Lightsail instance running a Linux-Nginx-MariaDB-PHP (LEMP) stack. PHP was likewise used to communicate with the back-end in real-time, which consisted of a Jupyter kernel and MariaDB database.

The machine learning algorithm was implemented in Python using the scikit-learn library (Pedregosa et al. 2011). Due to severe resource constraints on the Lightsail server, the algorithm was trained during the switch-over on a local machine, and the trained model was stored as a `joblib` binary then uploaded via `ssh` connection.

Interactivity between PHP and the algorithm was implemented using a modified Jupyter interface, which also handled queue management. Prediction response time during peak loads never exceeded 100ms.

The source code for the website is hosted on Github at `https://github.com/<REDACTED>`.

Open Source Attributions

This research would not have been possible without the enormous contributions of the open-source software community. In particular, the following libraries were integral to this research:

- `Scikit-Learn` (Pedregosa et al. 2011)
- `Jupyter` (Kluyver et al. 2016)
- `Numpy` (Harris et al. 2020)
- `Pandas` (team 2020)
- `Seaborn` (Waskom 2021)
- `Matplotlib` (Hunter 2007)
- `ggplot2` (Wickham 2016)
- `Tidyverse` (Wickham et al. 2019)

[8] CSS and other styling elements were copied then modified from `https://surveyjs.io/`.

Works Cited in Appendices

Ansolabehere, Stephen, and Shanto Iyengar. 1995. *Going Negative : How Attack Ads Shrink and Polarize the Electorate*. New York ; London: Free Press.

Ansolabehere, Stephen, Brian Schaffner, and Samantha Luks. 2020. "CCES Common Content, 2019." Harvard Dataverse. https://doi.org/10.7910/DVN/WOT7O8.

Baldwin-Philippi, Jessica. 2017. "The Myths of Data-Driven Campaigning." *Political Communication* 34 (4): 627–33.

Benjamini, Yoav, and Yosef Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society: Series B (Methodological)* 57 (1): 289–300.

Broockman, David, and Joshua Kalla. 2020. "When and Why Are Campaigns' Persuasive Effects Small? Evidence from the 2020 US Presidential Election." OSF Preprints. https://doi.org/10.31219/osf.io/m7326.

Coppock, Alexander, Seth J Hill, and Lynn Vavreck. 2020. "The Small Effects of Political Advertising Are Small Regardless of Context, Message, Sender, or Receiver: Evidence from 59 Real-Time Randomized Experiments." *Science Advances* 6 (36).

Dobber, Tom, Ronan Ó Fathaigh, and Frederik Zuiderveen Borgesius. 2019. "The Regulation of Online Political Micro-Targeting in Europe." *Internet Policy Review* 8 (4).

Goyal, Vineet, and Rajan Udwani. 2020. "Online Matching with Stochastic Rewards: Optimal Competitive Ratio via Path Based Formulation." In *Proceedings of the 21st ACM Conference on Economics and Computation*, 791. EC '20. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3391403.3399531.

Hainmueller, Jens. 2012. "Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies." *Political Analysis*, 25–46.

Harris, Charles R., K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, et al. 2020. "Array Programming with NumPy." *Nature* 585 (7825): 357–62. https://doi.org/10.1038/s41586-020-2649-2.

Hindman, Matthew. 2018. *The Internet Trap*. Princeton University Press. https://doi.org/10.1515/9780691184074.

Holm, Sture. 1979. "A Simple Sequentially Rejective Multiple Test Procedure." *Scandinavian Journal of Statistics*, 65–70.

Hunter, J. D. 2007. "Matplotlib: A 2d Graphics Environment." *Computing in Science & Engineering* 9 (3): 90–95. https://doi.org/10.1109/MCSE.2007.55.

Kluyver, Thomas, Benjamin Ragan-Kelley, Fernando Pérez, Brian Granger, Matthias Bussonnier, Jonathan Frederic, Kyle Kelley, et al. 2016. "Jupyter Notebooks - a Publishing Format for Reproducible Computational Workflows." In *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, edited by Fernando Loizides and Birgit Scmidt, 87–90. Netherlands: IOS Press. https://eprints.soton.ac.uk/403913/.

Krotoszynski, Ronald J., Jr. 2020. "Big Data and the Electoral Process in the United States: Constitutional Constraint and Limited Data Privacy Regulations." In *Big Data, Political Campaigning and the Law: Democracy and Privacy in the Age of Micro-Targeting*, 186–213. Routledge.

Kruikemeier, Sanne, Minem Sezgin, and Sophie C Boerman. 2016. "Political Microtargeting: Relationship Between Personalized Advertising on Facebook and Voters' Responses." *Cyberpsychology, Behavior, and Social Networking* 19 (6): 367–72.

Montgomery, Jacob M., and Santiago Olivella. 2018. "Tree-Based Models for Political Science Data." *American Journal of Political Science* 62 (3): 729–44. https://doi.org/10.1111/ajps.12361.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, et al. 2011. "Scikit-Learn: Machine Learning in Python." *Journal of Machine Learning Research* 12: 2825–30.

Sunstein, Cass Robert. 2015. "Fifty Shades of Manipulation." *Journal of Behavioral Marketing* 213 (1).

team, The pandas development. 2020. *Pandas-Dev/Pandas: Pandas* (version latest). Zenodo. https://doi.org/10.5281/zenodo.3509134.

Wachter, Sandra. 2017. "Privacy: Primus Inter Pares — Privacy as a Precondition for Self-Development, Personal Fulfilment and the Free Enjoyment of Fundamental Human Rights." *SSRN*, January.

Waskom, Michael L. 2021. "Seaborn: Statistical Data Visualization." *Journal of Open Source Software* 6 (60): 3021. https://doi.org/10.21105/joss.03021.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Zuboff, Shoshana. 2015. "Big Other: Surveillance Capitalism and the Prospects of an Information Civilization." *Journal of Information Technology* 30 (1): 75–89.