Persuasive Effects of Presidential Campaign Advertising: Results of 53 Real-time Experiments in 2016

Alexander Coppock
Yale University

Seth J. Hill
University of California San Diego†

Lynn Vavreck
University of California Los Angeles‡

August 23, 2019

Prepared for presentation at the 2019 meeting of the American Political Science Association in Washington, D.C.

Abstract: In this letter, we report the results of 53 randomized advertising experiments conducted over 29 weeks on 34,000 people during the US 2016 Presidential election. Our treatments were drawn in real time from advertisements on air each week. The ads vary on many dimensions: election type (primary or general), tone (attack or promotional), sponsor (candidates or Super PACs), context (timing), and content (topics). We manipulate which ads respondents see, when they see them, whether they see more than one ad, which ad they see first, and whether they see competing, reinforcing, or no additional information. Owing to the large size of our study, the meta-analytic estimates of the average treatment effects on favorability and vote choice are sometimes distinguishable from zero, but are always quite modest, even accounting for variation across advertisements and contexts.

*Assistant Professor, Department of of Political Science, 77 Prospect Street, New Haven, CT 06511; alex.coppock@yale.edu, http://www.alexandercoppock.com.
†Associate Professor, Department of Political Science, 9500 Gilman Drive #0521, La Jolla, CA 92093-0521; sjhill@ucsd.edu, http://www.sethjhill.com.
‡Marvin Hoffenberg Professor of American Politics and Public Policy, Professor, Departments of Political Science and Communication, P.O. Box 951472, Los Angeles, CA 90095-1472; vavreck@mac.com, http://www.lynnvavreck.com.
Persuasion is one of the central goals of every political campaign and in many races persuasion happens via televised advertising. In 2016, the two major-party nominees for U.S. president allocated 47.5% of their campaign expenditures to advertising.\(^1\) Despite increased attention on field organizing and mobilization, attempts to persuade voters through TV ads are thriving in campaigns for president.

Modest persuasive effects of presidential campaign ads and rapid decay has been demonstrated in both observational (Johnston, Hagen, and Jamieson, 2004; Hill et al., 2013; Sides and Vavreck, 2013; Sides, Tesler, and Vavreck, 2018) and experimental (Ansolabehere and Iyengar, 1995; Valentino, Hutchings, and White, 2002; Gerber et al., 2011; Bartels, 2014) research. Although it is common belief among campaign professionals that variations in attributes of advertisements affect persuasiveness (their tone, their sponsors, when they air), academic studies have typically been limited in their ability to explore whether the heterogeneity in advertisements corresponds with heterogeneity in persuasive effectiveness.

This is not to say that we know nothing about how the attributes of ads affect viewers. Prior experimental work has demonstrated differences in effectiveness at non-presidential levels due to the tone of the ad, the content, the sponsor, or the type of election (Brader, 2005; Weber, Dunaway, and Johnson, 2012; Brooks and Murov, 2012; Dowling and Wichowsky, 2015). Broader work on opinion change highlights the importance of context, competing or reinforcing information, and the recency of information on changing attitudes (e.g., Zaller, 1992; Druckman, 2004; Chong and Druckman, 2010) lending credence to the possibility that campaign ads vary in persuasiveness by things like who made them, what they say, when they are aired, and characteristics of the viewer. The professionals might be right, but a recent meta-analysis (Kalla and Broockman, 2018) of field experimental work on campaign persuasion concludes, “the best estimate for the persuasive effects of campaign contact and advertising . . . is zero.” Their study is quite valuable for learning about the effects of campaign contacts like mailers, door knocks, or telephone calls, but only one of the 49 field studies is of television advertising.

\(^1\)See https://www.opensecrets.org/pres16.
Our goal in this letter is not to introduce any new theory of campaign advertisements, but instead to update the empirical record on the possibly heterogeneous effectiveness of television advertisements for political persuasion. We conducted 53 survey experiments over 29 weeks on nationally representative samples totalling 34,000 people over 8 months of 2016 campaign. The ads we test are actual presidential advertisements resulting from strategies of presidential candidates trying to win the highest office in the land over the entire campaign. In this sense, our treatments — both content and timing — are determined by the equilibrium strategies of people who are highly motivated to persuade voters and win campaigns. This unique design (a) tests the ads that the best in the business thought would be effective and (b) maximizes external validity because these were the actual ads seen by voters delivered contemporaneously by the campaign.

Week after week, ad after ad, our experiments return modest effects. The estimated effects are consistently modest across many variations in tone, content, sponsor, election type, context, information environments, and characteristics of people. On some outcomes (respondents’ positions on the issues in the ads, the importance of the topics in the ads) we find effects that are so small they are essentially non-existent even though we can estimate them very precisely. On candidate favorability and vote choice, we find slightly larger effects that we can sometimes isolate from zero, but that look remarkably similar in magnitude across candidates and features of ads.

Research design

We ran our experiments from March to Election Day in 2016, covering the primary elections of both major parties and the general election. Each week, a representative sample of Americans were randomly assigned to watch campaign advertisements or a placebo ad and answer a short survey. Our design enables us to estimate effects of each ad separately and also to bolster the generalizability and the precision of our findings by pooling over all 53 experiments.

Subjects

Our subjects were recruited by YouGov, who returned to us samples of exactly 1,000 (or 2,000, depending on the week) complete responses. Some subjects (42%, on average) started the survey but
did not finish, which can cause bias away from our inferential target, the U.S. population average treatment effect (PATE). We rely on YouGov’s post-stratification weights to address the problem that some kinds of people are more likely to finish the survey than others. A second source of bias is the possibility that our treatments caused subjects to stop taking the survey. Using information on the full set of respondents who began each survey we find no evidence of differential attrition by treatment condition.\(^2\) Although this analysis does not prove missingness is independent of assignment, we proceed under the assumption that it is.

**Treatments**

We chose treatment ads from the set of ads released each week during the 2016 campaign by candidates, parties, or groups. We picked ads to test on the basis of real-time ad-buy data from Kantar Media and news coverage of each week’s most important ads. See the supplemental materials for more information, including transcripts, date of testing, and the ads themselves. We showed some ads in multiple weeks to estimate whether the effect of the same advertisement varies with changing context. In total, we tested 28 ads attacking Trump, nine attacking Clinton, nine promoting Clinton, and three promoting Trump. The remainder were promotional advertisements for other primary candidates.

**Random assignment**

The design of each experiment was very similar week-to-week. During the primaries (March to June 6th), 1,000 respondents were randomly assigned to watch either one of two campaign ads, both, or neither. Subjects who saw neither treatment ad were shown a placebo ad for Nationwide Insurance featuring Peyton Manning singing in his car. During the period surrounding the conventions (June 20th to August 15th), 2,000 respondents were recruited every other week. These subjects could be assigned to control or one of the two video conditions, but we removed the “both” condition during this period. In the fall, we returned to 1,000 respondents a week and re-introduced

\(^2\)Specifically, we conduct separate $\chi^2$ tests of the dependence between response and treatment assignment within each week of the study. Of the 29 tests, three return unadjusted $p$-values that are statistically significant. When we adjust for multiple comparisons using the Holm (1979) or Benjamini and Hochberg (1995) corrections, none of the tests remain significant.
the “both” condition on September 26th. We used Bernoulli random assignment to allocate subjects to treatment conditions with equal probabilities. On the basis of successful manipulation checks, we are confident that treatments were delivered as intended: tiny fractions of the control groups claimed to have seen campaign ads (2 to 4%) compared with overwhelming majorities of the treatment groups (92 to 95%).

**Outcomes**

After watching the videos, subjects took a brief survey, the full text of which is available in the supplemental materials. We asked respondents to rate candidates on five-point favorability, vote choice in the general and primary elections, and on nine issues candidates talked about in their ads. The issue items asked respondents to agree or disagree on four-point scales on terrorism, immigration, taxes, abortion, wages, guns, and discrimination. We also asked an open-ended question about the most important problem facing the nation.

**Estimation**

We estimate treatment effects separately for each week’s experiment using ordinary least squares (OLS) and robust standard errors. OLS is a consistent estimator of the ATE under our design (Lin, 2013). Exclusively in order to increase precision (Gerber and Green, 2012, Chapter 4), we control for the following covariates measured pre-treatment: seven-point party identification, five-point ideological self-placement, voter registration status, gender, age, race, income, education, region and a pre-treatment question about whether the country was on the right track.

The treatment effects in any single 1,000-person study are estimated with a fair amount of sampling variability but pooling across weeks with random-effects meta-analysis allows us to sharpen the estimates considerably. We chose random-effects over fixed-effects because we are uncomfortable asserting that the true effect of all ads is the same across all weeks of the campaign. Instead, we assume that the true effects vary week-to-week, but that they are drawn from a common distribution. The estimand for the meta analysis is the expectation (or average) of that distribution.
Main Results

In Figure 1, we present the average treatment effect estimates (and 95% confidence intervals) of advertising effectiveness across time (horizontal axis) and ad target (facet rows) on favorability for the target candidate. The second column of facets, labeled “meta-analysis,” is our random-effects estimate across all 29 weeks of experiments.

Figure 1: Effects of Ads on Target Candidate Five-Point Favorability.

Across weeks, sponsors, targets, and the content of ads, the effects of seeing one additional ad during an on-going competitive presidential campaign (relative to a placebo group) are quite small. We are, however, sometimes able to isolate these effects from zero. Substantively speaking, seeing one additional Clinton promotional ad increases a respondent’s rating of Clinton by 0.09 points (on a five point scale), with a standard error of 0.05 points. The effect size is similar for attacks on Trump. Seeing an ad that attacked Trump lowered respondents’ ratings of him by 0.07 points.
Effects for other types of ads (pro-Trump ads or anti-Clinton ads, along with other primary contenders) were of similar magnitudes, but because we tested fewer such ads, the meta-analytic estimates tend to be far less precise.

When we pool together all the attack ads, we find that on average, they decrease favorability of the target candidate by one-twentieth of a scale point. We can declare this very small estimate statistically significant because our meta-analytic standard error is one-fiftieth of a scale point. The pooled estimate for promotional ads is exactly the same magnitude in the other direction. Our sample of ads consists of more negative ads than positive ads, so the meta-analytic standard error is twice as big for promotional ads compared with attack ads, rendering the estimate statistically insignificant. This one percent shift in favorability (one-twentieth of a point on the five-point scale) is consistent with existing estimates of the effects of campaign advertising (see Gerber et al., 2011; Sides and Vavreck, 2013; Sides, Tesler, and Vavreck, 2018) and reinforces the conclusion that the effects of exposure to a campaign ad in the heart of a presidential year are small in size. Here we demonstrate that is true across different types of elections, periods of the campaign, sponsors and types of ads.

A natural concern is that the average effects shown in Figure 1 mask politically important differential response by different types of people. For example, Campbell et al. (1960) and authors subsequent have argued that individual attachment to party conditions the effects of any new political information. We present conditional average treatment effects (CATEs) separately for Democrats, Independents, and Republicans in Table 1 (See the supplementary materials for a graphical presentation analogous to Figure 1). Estimates are typically about a tenth of a scale point or smaller (for 26 of 33 coefficients), and the ones that are larger are usually estimated with less precision. To pick out just one row of the table, consider the effects of Anti-Trump ads. They are -0.11, -0.05, and -0.03 for Democrats, Independents, and Republicans, respectively. The effect is only statistically significant for Democrats, though the differences-in-CATEs are not statistically discernible from each other. Our interpretation is that the ads have approximately the same small effect across all three subgroups.
<table>
<thead>
<tr>
<th></th>
<th>Democrats</th>
<th>Independents</th>
<th>Republicans</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate (SE)</td>
<td>95% CI</td>
<td>Estimate (SE)</td>
</tr>
<tr>
<td>Anti Clinton</td>
<td>0.04 (0.06)</td>
<td>[-0.08, 0.16]</td>
<td>-0.08 (0.11)</td>
</tr>
<tr>
<td>Anti Trump</td>
<td>-0.11 (0.03)</td>
<td>[-0.17, -0.05]</td>
<td>-0.05 (0.09)</td>
</tr>
<tr>
<td>Pro Clinton</td>
<td>0.08 (0.09)</td>
<td>[-0.10, 0.26]</td>
<td>0.11 (0.19)</td>
</tr>
<tr>
<td>Pro Cruz</td>
<td>0.14 (0.33)</td>
<td>[-0.51, 0.80]</td>
<td>-0.26 (0.26)</td>
</tr>
<tr>
<td>Pro Kasich</td>
<td>-0.04 (0.19)</td>
<td>[-0.41, 0.34]</td>
<td>-0.55 (0.25)</td>
</tr>
<tr>
<td>Pro Sanders</td>
<td>0.16 (0.12)</td>
<td>[-0.08, 0.39]</td>
<td>-0.19 (0.22)</td>
</tr>
<tr>
<td>Pro Trump</td>
<td>-0.13 (0.17)</td>
<td>[-0.47, 0.20]</td>
<td>-0.08 (0.25)</td>
</tr>
<tr>
<td>Attack Ads</td>
<td>-0.08 (0.03)</td>
<td>[-0.13, -0.03]</td>
<td>-0.06 (0.07)</td>
</tr>
<tr>
<td>Promotional Ads</td>
<td>0.06 (0.06)</td>
<td>[-0.06, 0.18]</td>
<td>-0.06 (0.11)</td>
</tr>
</tbody>
</table>

Table 1: **Conditional Average Treatment Effects of Ads on Five-Point Candidate Favorability By Subject Partisanship.** Estimates were obtained via random-effects meta-analysis.

We also consider possible heterogeneities by two important dimensions for campaign tactics: the number of days until Election Day and whether or not subjects reside in “battleground” states. For both these analyses, we flip the sign of the attack ads, so that higher values always mean that ad moved candidate favorability in the intended direction. Averaging over all our experiments, the average effectiveness is 0.049 (SE: 0.02), or 1/20th of a scale point on a 5-point scale. We find that the slope with respect to time is about as flat as can be: effectiveness declines by about 0.0001 points per day (SE: 0.0003 points), or about 1/100th of a point every 100 days closer to Election Day. This decline is statistically insignificant. This result contrasts with a finding in Kalla and Broockman (2018) that campaign persuasion contacts appeared to be more successful the earlier they were deployed, but is consistent with previous findings about advertising persuasion (Gerber et al., 2011; Hill et al., 2013; Sides and Vavreck, 2013; Sides, Tesler, and Vavreck, 2018). Turning next to the contrast between battleground and nonbattleground states, we estimate that the effects are 0.003 points (SE: 0.04) higher in battleground states. This difference is substantively meaningless at 3/1000ths of a scale point and is statistically insignificant. Together with the analysis by subject partisanship, these heterogeneity analyses lead us to conclude that the small effects of ads are similarly sized for different groups of people across time and space.

Next we consider the effect of the advertisements on policy positions. We begin with subjects’ responses to the open-ended question that asked for their views on the “most important challenge
facing America these days.” Turning these open-ended responses into quantitative data involves a
number of analysis choices. We opted for a simple approach that reduced our own discretion. We
created ad-specific dictionaries based on transcripts of ads, then measured whether respondents
used one or more of the words in that dictionary in the open ended answer. On average, control
group subjects used at least one word in any of the treatment ads 20% of the time. On average,
treatment raised the probability of doing so by 1.9 percentage points (SE: 0.6 points). We interpret
this evidence to show that political advertisements can change the considerations that are most
salient when subjects think about political issues.

The ads may do more than change issue salience – they may change the way a voter thinks
about an issue. Among our 53 ads, 16 specifically tried to persuade voters on policies about
which we asked policy questions on our survey (the remainder were mainly about candidate traits).
Figure 2 shows the effect of these 16 ads on their target issue. Consistent with the overall theme
of our study, the meta-analysis suggests that ads about policies move people a bit in the intended
direction. Being exposed to a campaign ad about terrorism, immigration, the economy, or gun
ownership in 2016 moved people 0.06 scale points on a four-point scale in the direction the ad
intended (SE: 0.02 points).

Figure 3 and Table 2 show the effects of ads on general election vote choice. The estimates
range in size from zero to five points, some of which we can distinguish from zero. In some cases,
the estimates are opposite of what was intended. Attacks on Clinton, for example, seem to boost
her vote share by 1 point (SE: 1 point) and Trump’s promotional ads seem to lower his vote share
by five points (SE: 3 points), though neither of these “backwards” effect estimates is significant.
Attacks on Trump lower his vote share as intended by three points (SE: 1 point).

Pooling all the attack ads together we estimate the average effect of attacks to be a decrease
in target candidate’s vote share of about one point. Our estimate of this average effect is quite
precise, with a 95% confidence interval only 3 points wide that nevertheless contains zero. Our
understanding of the effects of positive ads is less sharp and we are unable to distinguish those
effects from zero.
Figure 2: Effects of Ads on Issue Positions. Measured with 4-point favor/oppose outcomes rescaled so that higher values indicate greater agreement with the position advocated by the advertisement.

In an analysis shown in the supplementary materials, we find that the effects on primary vote choice are similar in size. On average, promotional ads yield a 4-point boost in vote share (SE: 2 points) whereas attack ads return a 2-point decrease (SE: 3 points).

<table>
<thead>
<tr>
<th></th>
<th>Estimate (SE)</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anti Clinton</td>
<td>0.01 (0.01)</td>
<td>[-0.01, 0.04]</td>
</tr>
<tr>
<td>Anti Trump</td>
<td>-0.03 (0.01)</td>
<td>[-0.04, -0.01]</td>
</tr>
<tr>
<td>Pro Clinton</td>
<td>-0.00 (0.05)</td>
<td>[-0.10, 0.09]</td>
</tr>
<tr>
<td>Pro Trump</td>
<td>-0.05 (0.03)</td>
<td>[-0.12, 0.01]</td>
</tr>
<tr>
<td>Attack Ads</td>
<td>-0.01 (0.01)</td>
<td>[-0.03, 0.00]</td>
</tr>
<tr>
<td>Promotional Ads</td>
<td>-0.03 (0.03)</td>
<td>[-0.09, 0.03]</td>
</tr>
</tbody>
</table>

Table 2: Average Effects of Ads on General Election Vote Choice for Target Candidate. Estimates derived from random-effects meta-analysis.
Figure 3: Effects of Ads on General Election Vote Choice.
We turn now to a series of contextual features of the ad environment. First, we consider the effect of ordering, as it has been suggested that the order in which respondents are exposed to stimuli conditions effects on attitudes (e.g., Lohr, 2019). In 10 of our weeks, the group of subjects assigned to see two ads could see either the Trump or the Clinton attack ad first. Figure 4 plots the difference in favorability of the ad target depending on whether the attack was shown first or second. If recent ads are more accessible in memory, the estimates should be negative, since a more recent attack should decrease the target’s favorability more. As the figure makes plain, however, the order in which the advertisements were presented did not change their influence on candidate favorability.

![Figure 4: Effects of the Order of Ads on Favorability](image)

Next, we examine the eight weeks of the project in which we assigned subjects to see competing ads — for Clinton and Bernie Sanders in the Democratic primary election and Clinton and Trump in the general election. Figure 5 shows that seeing competing information from one’s opponent in no case changes the assessment of any of the 16 advertisements in these tests. Consider
panel (a), the effects of seeing a promotional ad for Sanders in the absence and presence of a competing advertisement for Clinton. The expectation is that the effect of Sanders promotional ad is neutralized by competing information from Clinton and vice versa. As the estimates in the difference row show, this expectation was not borne out. The same pattern holds for panels b, c, and d.

Finally, we turn to an analysis of eight ads that we fielded more than once, first in the week they were released by the campaigns and then again in later weeks. We did not have strong priors about how effects might vary over time. However, if ads are only effective when aired at a particular moment in a campaign, we learn both about the conditionality of persuasion and also about the external validity of our study. However, as Figure 6 shows, the effects of ads are remarkably consistent over time. In no case does the estimate of an ad’s effectiveness significantly vary with the week of experiment. For most of the ads, the effects bounce around close to zero. In the case of “Quotes,” the estimate appears to increase in size as Trump’s treatment of women increases in salience mid-year (the end of the primaries) perhaps in concert with a CNN interview on the topic, but this is entirely speculative.
Figure 5: Average Effects of Ads on Target Candidate Favorability, with and without Competition
Figure 6: Effects of Campaign Context on Ad Effects
Conclusion

On the basis of 53 experimental tests of presidential campaign advertising, we have demonstrated that the effects of exposure to an ad on candidate favorability, vote choice, issue salience and issue positions are small. The magnitude of the effect does not depend greatly upon characteristics of the ad like tone, sponsor, or target; characteristics of the information environment such as competing information, ordering of ads, or timing throughout the election year; or even descriptive comparisons across characteristics of people, like partisanship. In some cases when we pool our experiments, we have enough data to distinguish from zero the small average causal effects from campaign ads on favorability, vote choice, issue salience, and issue positions.

A central question unresolved in this study is whether our experimental effects imply advertising can have large effects in elections or not. On one hand, the effects we estimate are quite small. On the other hand, the treatments are one ad delivered in the heart of a campaign that aired hundreds of thousands of such ads. If effectiveness were to increase linearly in ads viewed (or if the marginal returns diminished slowly enough), our small effects could be consequential, consistent with the observed level of spending by candidates on media buys.

We close on a brief methodological note. While our meta-analytic estimates are quite close to zero, some of the individual experiments generate estimates that are large enough to be statistically significant. If we had run just one or two experiments and gotten lucky, we would have come away with an artificially inflated sense of the effectiveness of ads. In some cases, a single-shot experiment would have led us to conclude too-large effects in the opposite direction, what some have called “boomerang” effects. By self-replicating the same design more than fifty times, we guarded ourselves against these figments of sampling variability, which is especially important in substantive domains like persuasion, where effects tend to be small.
References


