How Campaign Ads Stimulate Political Interest

Nathan Canen and Gregory J. Martin

University of Houston and Stanford Graduate School of Business

ABSTRACT. We empirically investigate two key dynamic features of advertising competition in elections using a new dataset of very high-frequency, household-level television viewing matched to campaign advertising exposures. The resulting dataset comprises more than 100,000 ad airings. First, we show that exposure to campaign advertising increases households' consumption of news programming by 3-4 minutes on average over the next 24 hours. Our method compares households that were viewing a program at the instant a political ad appeared to other viewers in the same market who watched the same show but tuned out just before, or tuned in just after, the ad was aired. Hence, in contrast to many existing models, advertising indirectly influences voters' information environment by changing voters' decisions to become informed about the campaign. Second, we show that these effects decline over the campaign: the effect of political ads on news consumption is 30 percent lower close to election day than two months out, while the rate of tune-out of political ads increases towards election day. Together, these dynamic forces help to rationalize the observation that candidates deploy the bulk of their advertising budgets well in advance of election day.

KEY WORDS. Campaign advertising, electoral competition, dynamic effects.

JEL CLASSIFICATION: D72; D83; M37.

1. Introduction

The minimal effects hypothesis (Lazarsfeld et al., 1944; Berelson et al., 1954) suggests that campaigning has very small effects on voters' decisions. In this view, voters have persistent beliefs and do not change their voting or turnout decisions, even when exposed to new information from political campaigns. The early arguments for this position were based on observational evidence subject to the critique that one candidate's campaign effort is highly correlated with his/her opponent's, and thus the equilibrium marginal effect might be much lower than the true all-else-equal counterfactual. But modern field-experimental evidence (Coppock et al., 2020; Kalla and Broockman, 2018; Gerber et al., 2011), circumventing this problem through random assignment, has found effects on vote choice that are either very small or very short-lived. The Gerber et al. (2011) study, which randomized a portion of television ads by one candidate in a gubernatorial race, found effects on vote intention that decayed to zero within a week. Other studies find similarly small effects on turnout and voting behavior (Ashworth and Clinton, 2007; Krasno and Green, 2008 and most recently, Spenkuch and Toniatti, 2018 on turnout).

Politicians themselves do not appear to believe these findings. Campaigns devote enormous quantities of money and effort to advertising, debates, direct mail, door-to-door voter contact, and other forms of information transmission. Campaigns carefully target their advertising to programs whose audiences skew towards moderate swing voters, and voters likely to turn out (Lovett and Peress, 2015). And while effort and expenditures certainly ramp up towards the end of the campaign, candidates do not conserve their advertising budgets for an immediate pre-election-day blowout, as would be suggested by a straightforward reading of the experimental evidence: in our data, more than 80% of general election TV advertising impressions occur two or more weeks prior to election day.

Are these campaign efforts wasteful, from the candidate's perspective? We propose and empirically investigate a new mechanism that can rationalize early campaign advertising expenditures, even if the direct changes to turnout or vote intentions they induce are minimal or short-lived. We study whether exposure to campaigns spurs political interest, inducing voters

to inform themselves by consuming more political news.[1] Hence, even if ads have minimal immediate effect on voter preferences, advertising can still change voters' information sets at election time by changing if and when voters become informed about the election. This opportunity declines as the election approaches, because there are comparatively fewer uninformed voters whose beliefs can be moved by new information.

Our explanation focuses on an under-studied aspect of campaigns: their dynamic nature. Both politicians and voters send and receive information on multiple occasions over the course of a campaign. A single TV ad or other voter contact in isolation may not have much direct impact on voter beliefs or actions, as the experimental evidence generally confirms. But voters' total accumulation of political information over an entire campaign — from campaign activities as well as from media coverage — dwarfs that of any single ad. Indeed, there is a large body of evidence that quasi-random variation in the macro-scale information environment has substantial impacts on voter behavior.[2] We connect these two disconnected bodies of evidence by investigating whether an advertising intervention at time t can affect voting at time T by changing the voter's information acquisition decisions at $t+1, t+2, ..., T$.

To test this mechanism, we implement a difference-in-differences design exploiting variation in exposure to ads using very high-frequency set-top box data. We are able to observe television viewership by more than 200,000 households at 1 second intervals. We compare a treatment group of viewers who were watching a program at the instant a political ad began with a control group of viewers that tuned out from this same program slightly earlier, or tuned into it slightly later (and hence, did not see the ad). Every political ad within our

---

[1]Mechanisms underlying this stimulus might include: (i) the ad makes a factual claim, which induces viewers to seek out corroboration from a more neutral source, (ii) the quantity of ads is an indicator of the intensity of political competition or likely closeness of the race, increasing interest in the campaign, and/or (iii) the ad makes a viewer aware of a candidate they didn't know existed, stimulating interest in learning about that candidate. We revisit these mechanisms in Section 4.

[2]Such informational variation could take the form of exposure to campaign advertising, as in Gordon and Hartmann (2013) or Spenkuch and Toniatti (2018); exposure to partisan media as in DellaVigna and Kaplan (2007) and Martin and Yurukoglu (2017); release of newspaper endorsements (Chiang and Knight, 2011); or release of evidence of corruption (Ferraz and Finan, 2008). Large information campaigns have also been shown to impact voters' beliefs in developing countries (Cruz et al., 2020). Our point is echoed in recent work by Le Pennec and Pons (2019), who find minimal effects of single TV debates on vote shares across many OECD countries, even though there is evidence that information in general does affect outcomes. As a result, they conclude that it is the persistent exposure to information that generates effects on electoral outcomes.

sample is a quasi-experiment inducing distinct treatment and control groups, generating over 100,000 such quasi-experiments. Our results show an average net increase of approximately 3-4 minutes (about a 5-6% increase compared to baseline viewing in the treated group) in news viewing over the next 24 hours following exposure to an ad. This effect is decreasing as election day approaches, from over 5 minutes early on in the campaign, to under 3 minutes close to election day. The fact that political ads frequently run on news programs[3] generates a positive feedback, as ad viewers are more likely to encounter additional political ads in the future. This feedback loop keeps viewers exposed to political information in both the ads themselves and in surrounding news content. As a consequence, campaigns appear to be able to stimulate increased attention to politics through their campaigns, affecting both political interest and engagement in campaigns by citizens. This finding contrasts with results that interpret the lack of effects of advertisements on turnout to mean that campaigns do not spur political engagement (Huber and Arceneaux, 2007).

Our effects are heterogeneous, and vary according to both the type of race for which the ad was aired and the identity of the sponsor. The effects are strongest for ads in gubernatorial races, weakest for ads by outside groups, and are stronger for ads run by challenger candidates than those run by incumbents. The effect appears to be driven by increased viewership of local news (rather than the national cable channels). These results together provide evidence for the mechanism of ads inducing voters to seek out information. Voters are more likely to recognize and already possess information about incumbents than challengers. And except for the highest-profile governors, local news is a more likely source to find coverage of state-level politicians. This interpretation is supported by heterogeneity analysis on the viewers' side. The effects are stronger for viewers who otherwise consume less political information: those on the lower end of the distribution for age, income and education.

We validate our results with a Regression Discontinuity Design (RDD) using tune-in time as the running variable. This design compares households who "barely" missed watching an ad, tuning in just after it aired, to those who "barely" saw it, tuning in just before it aired. While

---

[3]About 40% of political ads in our sample occurred on a news program.

the RDD requires pooling ads into a single sample for statistical power,[4] we show that there is a positive and significant jump of over 2 minutes of news consumption in the following 24 hours. The RDD also allows for a natural placebo test: we replace the outcome with news viewing in the *previous* 24 hours. We find a very precise zero effect in this placebo specification.

We probe our main results with a number of robustness tests. Among them, we show that there are no effect of political ads on non-news viewership (e.g. there is no increase to households' average viewing of sports or comedy programs). We also perform a bounding exercise inspired by works on partial identification in reduced form settings (e.g. Flores and Flores-Lagunes, 2013), where the bounds are constructed from control groups most likely to violate the identification assumptions.[5] We find that the estimated set defined by these bounds is reasonably tight, positive and statistically significant.

We also test how viewer attention to ads themselves varies over the course of the campaign, and with viewer attributes. We measure, among the set of viewers tuned in at the instant a political ad begins, the fraction who change the channel before the end of the ad. As a baseline for comparison, we compute the viewer-specific tune-out rate at another, randomly selected, time during the same day. We find that, as the election approaches, the tune-out rate increases at a rate of 0.02 p.p. per day (the average rate is 4.2%). This result is consistent with the mechanism of viewer interest in the campaign being stimulated by political advertising, and also strengthens candidates' incentives to advertise early, before voters are saturated with advertising. Nevertheless, the low absolute tune-out rates mean that "noncompliance" with treatment in our setting is low: viewers who are viewing when a political ad begins typically sit through it.

Our results connect and contribute to several strands of research. First, our results add subtlety to standard Bayesian voting models used in formal work on campaigning (e.g. Achen,

---

[4]In the RDD, the window defining treatment and control subjects can be as small as only a few seconds, compared to the few minutes in the difference-in-differences case. This means per-ad sample sizes are much smaller.

[5]The lower bound is derived from households who tune-in just after the ad is aired and are likely to be increasing their news consumption, while the upper bound is derived from those who tuned out just before the ad and are likely to be reducing their news consumption.

1992; Gerber and Green, 1998). If ads can be treated as (partially) informative signals and voters are Bayesian, willingness to search for information should decrease with every additional ad a viewer sees.[6] This is the opposite of the stimulating effect we observe in the data. As a result, we provide evidence that can inform models of information acquisition and accumulation. Instead of viewing ads as containing information themselves, our results suggest a model of ads as increasing viewer interest in campaign information from other sources. Or, equivalently, reducing viewers' search costs for campaign-related information. This mechanism helps to explain why, for example, the randomized information campaign of Kendall et al. (2015), which provided information to voters exclusively about the incumbent mayor's record and positions, reduced the uncertainty of treated voters' beliefs about his *opponent*; or why Gerber et al. (2009) find that the randomized delivery of the Washington Times, a conservative newspaper, increased the support for Democratic candidates in 2006.

Ads are thus a means for politicians to modulate the parameters of voters' information search problem. As detailed in the theoretical results of Matějka and Tabellini (2020), these parameters have important implications for politicians' platform choices and equilibrium policies. The interest-stimulating effect gives politicians some control over the timing of voters' information acquisition, a strategic problem modeled by Gratton et al. (2017).

Second, we provide new results on the effects of television advertising in campaigns. Our very granular and high-dimensional data provides advantages to those mostly used in the literature,[7] as we can explore variation in exposure due to very fine variation in the timing of TV viewing across citizens, while conditioning on media market, program characteristics and viewer preferences. This variation is very rich, generating many quasi-natural experiments. The scale of the data allows us to evaluate heterogeneity along several dimensions, including the effects of advertisements by different types of candidates, during different parts of the campaign, and on different subsets of viewers.

---

[6]As priors get tighter, the benefit of new information decreases. If costs of information acquisition are constant, then search should decline. This intuition also holds in a strategic setting where politicians can target the timing of ads to citizens (Gratton et al., 2017).

[7]As Kalla and Broockman (2018) describe, "First, the existing literature (and, by extension, our meta-analysis) provides only scarce evidence on the effects of television and digital advertising, which represent a great deal of campaign spending...more evidence about these mediums would clearly be welcome." (p.163)

When compared to the existing literature on television advertisements and turnout or vote shares (e.g. Ashworth and Clinton, 2007; Krasno and Green, 2008; Spenkuch and Toniatti, 2018), we differ in focusing on how campaigns change viewers' consumption of political information, rather than voting behavior. While there is some literature on interactions between paid and "earned" media (Ridout and Smith, 2008; Lovett and Staelin, 2016), and on the influence of campaigning on the media environment more broadly (Vavreck, 2009), we provide direct evidence of this linkage, and a mechanism that works through induced changes in viewer demand for campaign information. Instead of focusing on the static effects of an ad on viewers' contemporaneous preferences — which may well be minimal — we examine their dynamic influence on the complete informational picture that voters absorb. Our results imply that political advertising has a sizable informational externality, which is both relevant for the regulatory treatment of advertising and likely to influence equilibrium political competition.[8]

While our focus is on the consumption of political information rather than voting behavior, the two are clearly linked, and our results have implications for the likely magnitude of advertising effects on election outcomes. While we do not directly measure this link, we can use existing estimates in the literature on persuasive effects of media to roughly estimate the possible magnitude of the implied effects on vote shares. Our back-of-the-envelope calculations suggest that the news-consumption-stimulating mechanism could explain a significant fraction of the effects found in Spenkuch and Toniatti (2018), for example.[9]

Finally, our methodological approach could also be useful to other researchers with similar dynamic and high-dimensional settings. Our implementation differs from the standard difference-in-differences design because we face households that can be subject to multiple treatments over multiple periods, and they can switch treatment status over time (i.e. those

---

[8]For example, see Avis et al. (2020) and Fouirnaies (2018) for evidence that increases in campaign expenditure caps encourage greater entry by wealthy candidates. Analogously, advertising-driven changes in which kinds of voters are likely to be exposed to campaign information could change which kinds of candidates are likely to run.

[9]The confidence bands on this extrapolation exercise are wide, but our upper bound estimate includes the observed effects in Spenkuch and Toniatti (2018). We provide more discussion on this exercise and its limitations in Section 4.

who barely tuned in to ad a might have barely tuned out of ad $a' \neq a$). Aggregating multiple experiments is known to suffer from multiple statistical shortcomings (e.g. Abraham and Sun, 2020; de Chaisemartin and d'Haultfoeuille, 2020), while empirical designs with changing treatment statuses are still an active subject of research.[10] We address these issues by estimating the model at the ad-level. This generates ad-specific effects that are valid under the usual identification conditions for difference-in-differences designs. This approach has the additional advantages of computational simplicity in estimation and it yields distributions of ad-specific effects which can be used for heterogeneity analysis. However, pursuing standard identification tests to validate this design (e.g. pre-trends and balance checks) is nontrivial: one cannot simultaneously view the over 100,000 pre-trend graphs, for instance. We proceed with distributional approaches that leverage this dimensionality instead. For instance, we compare the distribution of p-values for the null of no difference between treated and control groups to the uniform distribution which would hold if the null is true to check for balance on observables. Finally, we discuss the additional requirements for such dynamic designs to be valid. One of them is the requirement that that treatment/control groups in ad a are balanced in terms of *past* treatments. We show this holds in our setting. All of these tests collapse to standard checks in the one treatment-one period model.

2. Data

We use two primary data sets in our analysis: household-level television viewership from set-top boxes, and ad occurrences. The former comes from the vendor FourthWall Media; the latter comes from Nielsen. We briefly describe each dataset in turn.

2.1. Viewing Data

Our viewership data covers the subscribers to 10 cable providers (Multiple System Operators or MSO's) around the country. The dataset covers viewing by more than 200,000

---

[10]There has been recent interest in generalizing these difference-in-differences designs, but we are not aware of methods that can easily accommodate our set-up due to its dimensionality or the identification of our parameters of interest. We revisit this discussion in Section 3.

FIGURE 1. Distribution of DMA's in our final sample
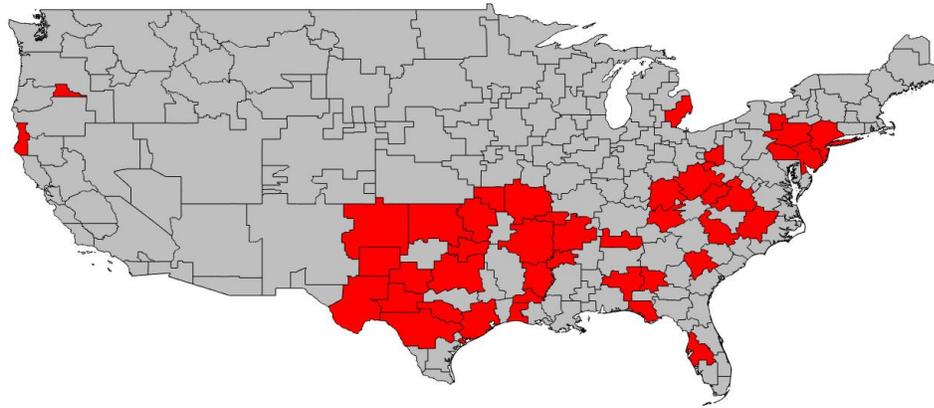
Notes: The figure presents the geographical location of DMA's in our final sample of ads. Our data is more concentrated in smaller cities and more Republican leaning regions. The names of the DMA's are presented in Table A.1 in Appendix A.

anonymized households from June 2012 to January 2013,[11] spanning 60 Designated Market Areas (DMAs).[12] Importantly, the data are not an opt-in sample, but cover the population of subscribers to each of the 10 MSO's. Compared to the general population, our sample is distributed more in smaller cities and further East and South than the average viewer, as shown in Figure 1.

The data are event-level, tracking tuning decisions (e.g. changing the channel or turning off the set-top box) to a resolution of 1 second, for each set-top box or "device." We observe the time of a tuning event and the channel switched to. We use the event data to construct viewing intervals: the time span between two tuning events.[13] The viewing interval data then allow us to determine the set of devices exposed to a given political ad, using a method detailed below.

---

[11]The number of households covered varies by day, but the median day in our sample period has 234,834 active households.

[12]While these MSOs have some subscribers in 60 DMAs, they are concentrated in a smaller subset. Systems are usually defined at municipality level, and hence it is possible for an MSO to operate only one or a few systems comprising a very small fraction of the population of some DMA. This is the case for many of our providers, who often have scattered subscribers in many DMAs but the bulk of their subscribers in one or two. The top 10 DMAs in our data account for more than half of subscribers.

[13]Figure A.1 in Appendix A provides a visualization of the tuning activity for an example device on one day.

The outcome of interest is viewers' consumption of news programming, which can be measured in the same set-top box data. We use a database of program classifications provided by FourthWall to determine which programs qualify as news and which do not. Specifically, we include any programs which FourthWall tags as in the "News" or "Politics" genres, except for any program which also includes any of "Entertainment", "Sports non-event", "Sports event", or "Religious" in the genre field. The latter is to exclude programs like ESPN SportsCenter or Access Hollywood; our interest is in measuring viewers' consumption of politically-relevant news. We additionally count any program airing on the 24-hour news channels CNN, Al-Jazeera English, MSNBC, Fox News Channel, and CSPAN as news. We use this classification to determine which viewing intervals corresponded to news programs, and measure consumption as the total minutes spent in viewing intervals corresponding to news programs.[14] News viewing outcomes are measured at the device level.

The set-top box data also provides a rich set of demographic covariates. Further detail on the set of covariates in the data is provided in Table A.2 in Appendix A. Covariates are measured at the household level.[15] As the majority of households contain more than one individual, covariates that are measured at individual level (such as age, or race) are the values for the head of household.[16]

---

[14]The Fourthwall program database is incomplete and has some missingness, particularly for local news programs. For example, in the Bend media market (where one of the MSOs covered by our data is located), most days in the sample period have no local news programs appearing in the program database. The cable news program schedule is much more consistent and complete but also has some idiosyncratic missingness. We note here that this missingness affects the efficiency of our estimates but not consistency (as long as the share of missing data vanishes with sample size) and cannot generate a bias away from zero, as program schedules are shared across all viewers in a DMA and thus the measurement error it induces cannot be correlated with treatment group status. Appendix D presents results where we drop ads in markets with the most missing news program data, which are qualitatively similar to the main estimates.

[15]A household can have more than one device (set-top box), although the median household in our data has one. 54% of households in the data have one device, 29% have two, 10% have three, and the remaining 8% have four or more. Outcomes will be measured at the device level. We do not aggregate to household level because the number of devices is strongly and positively correlated to measures of household size, and aggregating thus increases the likelihood of pooling viewership and treatment status across multiple individuals.

[16]Table A.3 in Appendix A compares mean values of demographic variables in our sample to those in a contemporaneous nationally representative sample of cable TV subscribers, the GfK/MediaMark Survey of the American Consumer. Our sample is whiter, lower income, less likely to be college educated, more likely to be homeowners but with lower housing and other wealth, than the national average. Some of these differences are at least partially attributable to measurement differences. In particular, continuous variables like income and wealth are binned and top-coded differently, and partisanship is measured differently in GfK compared to our data. See the notes to Table A.3 for details.

2.2. Ads Data

Data on political advertising comes from Nielsen's Ad Intel database.[17] We capture the universe of political advertisements aired in the 2012 cycle on spot TV.[18] The Nielsen data includes the DMA in which the ad ran, the date and exact air time, the program and network on which the ad ran, the sponsor, the impressions and Gross Ratings Points (GRPs), and a descriptive title. Data are at the level of the individual ad occurrence. Using the sponsor information, we determined the type of sponsor (candidate or outside group) and, for candidates, the office sought. We limit to ads run in the 60 DMAs covered by the viewership data. The resulting dataset has 286,863 individual ad occurrences, the vast majority of which (96%) are standard 30-second length; the remainder are split approximately equally between 15 and 60 second lengths. In Figure A.2 in Appendix A, it can be seen that over 80% of expenditures occur more than 2 weeks prior to election day.

Because many DMAs in the set of 60 covered by the set-top box data have only a small number of covered households, many of these ads have no households in our sample in either treatment or control groups. Dropping these ads with no viewing data reduces the number of ads in the final dataset to 112,480. Table 1 shows the distribution of ads with nonzero set-top box households in both treatment and control groups, by sponsor.

2.3. Linking Viewership to Ad Occurrences

We consider a set-top box in the viewing data to be exposed to a given political ad if its viewing history meets the following criteria. First, the box was tuned to the channel on which the ad ran in the second at which the ad *began*.[19] Second, that the box had tuned to the

---

[17]Nielsen data is a commercial product and requires subscription or institutional access. An alternative data source is the Wesleyan Media Project, http://wesmedia.wesleyan.edu, which is available to qualified researchers for a nominal fee. The datasets are not identical, but the overlap is quite high ($> 80\%$) with our data source.

[18]Ads in the sample ran on local network affiliates, each of which reach a single media market. The dataset does not cover national ad buys (which are rare in presidential campaigns and nonexistent in down-ballot races) or local cable buys.

[19]I.e., ours is an intent-to-treat estimate. It is possible that there is noncompliance, if viewers in the treated group tune out before the end of the ad. Section 5 provides evidence on the rate of tune-out among viewers in the treated group.

TABLE 1. Distribution of Ads by Sponsor used in our Empirical Strategy

Sponsor Type	Ad Count
US House	24664
US Senate	15438
President	34257
Statewide	12954
Outside Group	25167

Notes: We present the ad count by sponsor in the Nielsen data, conditional on having at least one device in both treatment and control groups. The first four rows are candidate-sponsored ads, grouped by office sought. The last row are ads sponsored by third-party independent expenditure groups.

channel on which the ad ran no more than twenty minutes prior to the ad start time. This second restriction is in place to remove viewers who leave their television on for long stretches but may not be actively watching. The set of set-top boxes which meet these two criteria are the treatment group for a given ad.

We define a control group of set-top boxes at the ad level using the following criteria. First, the box was active (not off) at the second in which the ad began, but was not tuned to the channel on which the ad ran at that time. Second, the box must have been tuned to the same program and channel on which the ad ran at some point during the same half-hour time block in which the ad ran. Third, it began viewing the channel on which the ad aired no more than twenty minutes prior to the ad start time.[20]

These conditions are necessary to generate a valid control group for the treated devices. The second condition is the most fundamental, and deals with the issue that campaigns target advertising purchases to specific programs on the basis of desired political characteristics of the program's audience (Lovett and Peress, 2015). Because targeting occurs at the program–time block–media market level, constructing a control group that also viewed the same program in the same time block and the same media market as the treated group ensures that treated and control groups do not differ on unobservables known to advertising buyers.

The first and third conditions are needed because we impose the same conditions on the treatment group. We require that a "treated" device is on at the time of ad airing, for obvious

---

[20]In Appendix C.3, we show that our point estimates are very similar if we consider a five-minute or a ten-minute block for active set-top boxes group instead of the twenty-minute one.

FIGURE 2. Distribution of treatment and control group size (number of devices)
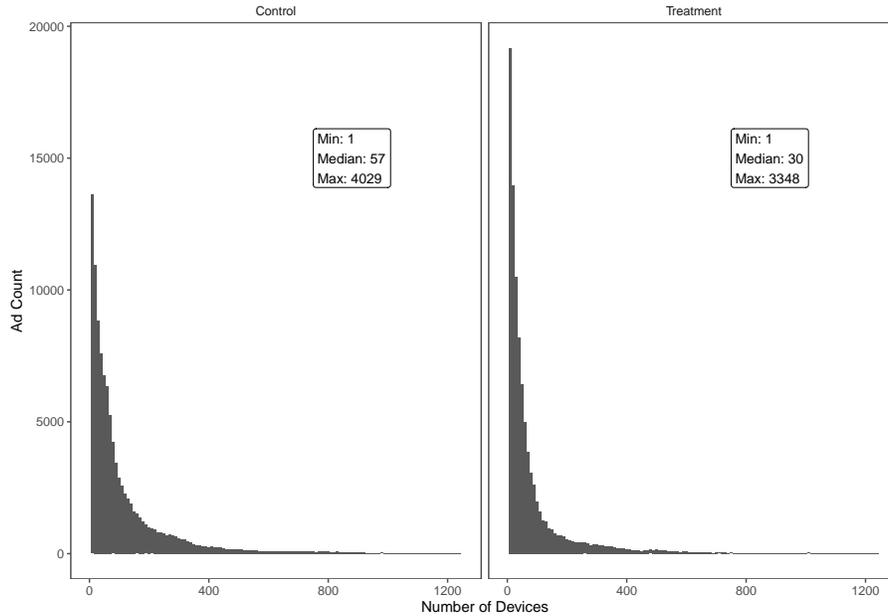
Notes: We present the distribution of the group size in number of devices for treatment and control groups. Figure excludes 383 groups with size exceeding 1250 devices.

reasons, and had begun watching in the past twenty minutes, in order to increase the likelihood that the viewer was actively watching at the time of ad airing. If we did not also impose the same conditions on control devices, we would put the parallel trends assumption at risk, a condition for identification of our model which we expand on in the next section.

The typical size of treatment and control groups in the sample is in the range of 30-60 devices. Figure 2 shows the full distribution of group sizes in the sample across all 112,480 ads.

3. Empirical Strategy

Our aim is to study the effects of exposure to political ads on news consumption. Consider a time window w relative to ad a, where $w = 0$ indicates the window prior to air time and $w = 1$ indicates the symmetric window following air time. We will use a window length of 24 hours in our analysis, as this allows us to difference out daily cyclical variation in news

viewing.[21] Let $y_{a,i,w}$ denote the total amount of time that device i spends watching the news in time window w.

Device i can be in either the treatment or control groups, with treatment status indicated by the random variable $T_{i,a}$. $T_{i,a}$ equals 1 in case of treatment (device was viewing the program of ad airing at time of ad airing) and 0 in case of control (device viewed the same program but was viewing something else at the time of ad airing).

When treatment effects are possibly heterogeneous across ads, the model is:

$$(3.1) \qquad\qquad y_{a,i,w} = \alpha_a + \beta_a T_{i,a} + \gamma_a w + \delta_a T_{i,a} w + \varepsilon_{i,a,w}.$$

with δ_a being our parameter of interest for each ad, a. δ_a captures the average effect of being exposed unexpectedly to political ad a on news consumption in the following 24 hours (relative to the average change in news consumption among those not exposed). In a design with homogeneous effects, we would simply have $\delta_a = \delta$ for all ads. In a standard design with only one treatment, α_a and β_a could be simply denoted as α and β.

The difference-in-differences design is important as it allows us to eliminate two sources of bias.[22] First, the possible existence of secular time trends in viewing. Such trends are likely in our case because being in either treated or control group requires actively watching television, and there is thus a mechanical increase in the expected future news viewing relative to the past, where the TV may have been off. The treatment-control difference eliminates this mechanical increase. Second, possible differences in viewing behavior between the types of viewers who are watching during an ad block and those who tuned out earlier or tuned in later. It is conceivable that the former type watches for longer stretches at a time and thus will have higher measured viewing in any given time window. The within-group before-after difference eliminates this bias.

The remaining threat to identification is eliminated if the parallel trends assumption is satisfied: that both treatment and control would have had the same *change* in news viewership

---

[21]Appendix B.1 gives a formal argument for the choice of 24 hours.

[22]As mentioned earlier, our use of within program-time block-media market variation additionally eliminates selection bias due to the targeting of advertising to specific viewer groups.

absent exposure to the ad. Parallel trends is fundamentally untestable, as we cannot observe counterfactual viewership for the treated group. However, we have available a rich set of viewer demographic characteristics and can check that treated and control groups do not exhibit substantive differences on observable dimensions. And, we can compare viewership trends in the pre-exposure period, to check that any divergence between groups appears only after the ad time. We show the results of these tests in the next subsection. There, we provide further supporting evidence on the validity of the parallel trends assumption in our setting.

By allowing heterogeneous effects, we recover a vector of parameters of interest, $\{\delta_a\}_a$, which can vary across subsamples — including when the ad is aired during a day (for example, the effects of ads aired early in the morning could differ from those of ads aired in the evening), the timing of the ad in the campaign (whether in the beginning or in the last few days), or the type of election (e.g. presidential vs. congressional).[23] Our results in Section 4 illustrate that it is important to allow for such heterogeneity.

We pursue identification and estimation of equation (3.1). Identification assumptions here are relaxed relative to a model with homogeneous effects $\delta_a = \delta$ for all ads, in that we require parallel trends to hold only within-ad and not across treatment (T) and control (C) groups from different ads. As a result, the conditions described under equation (3.1), which are common to all difference-in-differences designs, are not strictly necessary for the validity of the design. This relaxation will prove to be important due to the cyclical nature of news viewing over the course of the day, and variation in audience composition across programs and time blocks. In the next subsection, we report *distributions* of test statistics from tests of balance and pre-trends across all ads in the sample, also necessary for the validity of the design with heterogeneous treatment effects.

---

[23]This contrasts to the standard difference-in-differences design with an individual ad. This wealth of heterogeneity implies further computational and identification challenges due to the over 100,000 different "experiments" that must satisfy the identification assumptions in this design. We address this in the next section.

3.1. Exogeneity of Treatment

3.1.1. **Balance Tests.** We now show that the covariates of treated and control groups are balanced, an important check for the validity of our design. Due to the high dimensionality of our data with multiple experiments, we must find ways to appropriately present such balance tests. We provide two alternatives: the first, shown in Figure B.1 in Appendix B, presents the distribution of the estimated "treatment" effects for each covariate across ads. Each observation underlying the reported kernel density estimates is the regression coefficient of some predetermined covariate (such as age or income) on treatment or control status, for a single ad. We report a distribution instead of a point estimate because we have many ads. Under the null that groups are balanced, this distribution should approximate a normal distribution centered at zero. An alternative is to plot the (distribution of) p-values across ads for the treatment effects on each covariate, rather than the coefficients themselves. Under the null hypothesis of no differences across treatment and control groups, we expect a uniform distribution of p-values. We report these distributions in Figure B.2 in Appendix B.

Our covariates appear balanced across treatment and control groups. This can be clearly seen in both sets of graphs, where treatment effects have an estimated distribution precisely centered around 0 for all measured covariates. Given the large number of ads, it is natural that some would be drawn with statistically significant differences across characteristics, but the central tendency is very close to zero for all measured covariates. The distribution of p-values for treatment effects across ads (in Figure B.2) closely follows the theoretical uniform distribution as well.[24]

One final balance check is particular to our set-up with many "events". Our parameters of interest, δ_a, represent the average effect of exposure to the ad a on future news viewership. However, campaigns are dynamic: viewers are exposed to multiple ads along the campaign. To guarantee that we identify the effect of the current ad a instead of a history of treatments,

---

[24]While there are a few possible exceptions in Figure B.2 (e.g. `Hispanic`), Figure B.1 suggests that these low p-values are driven by the very large sample size and corresponding precision of our estimates. Our exercises in Appendix C.3 show that our results are qualitatively similar (and in fact increase in magnitude) when we drop the subset of ads with statistically significant ($p < 0.05$) differences in those covariates.

treatment and control groups should be balanced in their exposure to *past* ads.[25] For example, consider the extreme case that ads are always watched by the same viewers, while the other viewers are always "control". Then our specification would be comparing viewers with multiple treatments to those with none — failing to identify the effect of a marginal ad. Figure B.3 in Appendix B shows that this is not the case: there is substantial within-device mixing of treatment assignment, with most mass close to the 45 degree line. The typical device in our sample appears about equally often in both treatment and control groups. This validates our interpretation that δ_a identifies the effect of ad a alone, and not the compiled effect from multiple past ads.

As a result, our experimental design has control and treatment groups that appear well balanced across both observable characteristics, as well as on the frequency of treatment.

3.1.2. **Parallel Trends.** In the standard difference-in-differences (DiD) model, checking for the absence of pre-trends simply involves comparing trends in the outcome variable in the pre-treatment period across two groups. While the standard DiD model has two periods (pre and post-treatment) and two groups (treatment and control), we are faced with multiple periods (one for each ad over multiple days); multiple dosages (i.e. the same subjects can be treated multiple times), and heterogeneous treatment effects (ads aired at different times or by different sponsors may have different effects).

One way to visualize possible pre-trends is to aggregate all ads and plot the average news consumption for treatment and control groups across windows before and after ad exposure. This is the standard graphical representation in event-study designs: normalize all treatments to a period "0", and compare outcomes before and after that period across groups. We present this first in Figure 3a, which shows that control group viewing closely tracks that in the treatment group prior to ad exposure.

However, as discussed in Abraham and Sun (2020), this test might be inappropriate in the presence of heterogeneous treatment effects. For example, most people watch news in the

---

[25]We will study how the increase in news viewership due to an ad affects exposure to future ads in Section 4. However, this *ex-post* outcome is different to the treated group having more ads *ex-ante*.

morning and/or in the evening, so the time-pattern of viewing in a 24 hour window around 5am will look very different than that in a 24 hour window around 8pm. It follows that aggregating these effects might assign spurious pre-trends: the pre-exposure control group at 5am is not an appropriate comparison group for the pre-exposure treatment group at 8pm. Under different treatment effects, the weight attributed to those at 5am conflated with their differential trend can show up as an inappropriate pre-treatment trend break. In fact, aggregating heterogeneous effects across time and treatments is known to generate multiple problems, as the aggregation weights can be inappropriate (Goodman-Bacon, 2018; de Chaisemartin and d'Haultfoeuille, 2020, among others).

A solution in our context is to appropriately control for the "news viewing cycle" and the sources of heterogeneity in treatment effects. Most notably, these could be variations within a day and media markets, as well as the day, hour, and timezone an ad is aired, for both treatment and control groups. Ideally, we would visualize such trends at the ad-level (the level upon which this assumption is made), or at least control flexibly for such confounding factors. Given we cannot present graphs or results for all experiments, we present the results from plotting residuals for hourly news viewing before and after the ad, saturating for fixed effects at the time-zone by hour-of-day level. This is shown in Figure 3b below.

The results suggest the lack of systematic differential pre-trends across treatment and control groups. We note that the increase in news viewership for both groups before treatment in the graph is due to how we define treatment itself: devices must be active at the time of ad airing to be included in either group, implying viewership will be higher than average over the whole 48-hour period, during some or most of which the device may be off.

Figures 3a and 3b also visually display the cyclical pattern in news viewing, which for both groups rises around the time 24 hours before and 24 hours after the ad time. This pattern is reflective of viewers' habits: many viewers regularly watch the same program at the same time each day. In Appendix B.1, we present a formal argument showing that our 24-hour-window estimate is robust to differential trends between treatment and control groups (i.e., violations of the parallel trends assumption) if those differential trends have a 24 hour cycle.

FIGURE 3. Pre-Trends
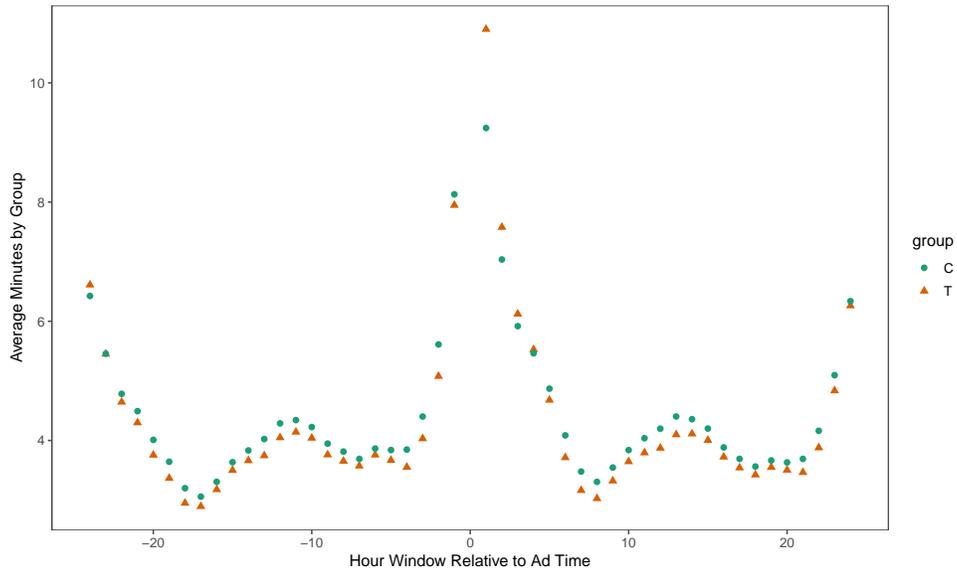
(A) Hourly average news viewing across groups
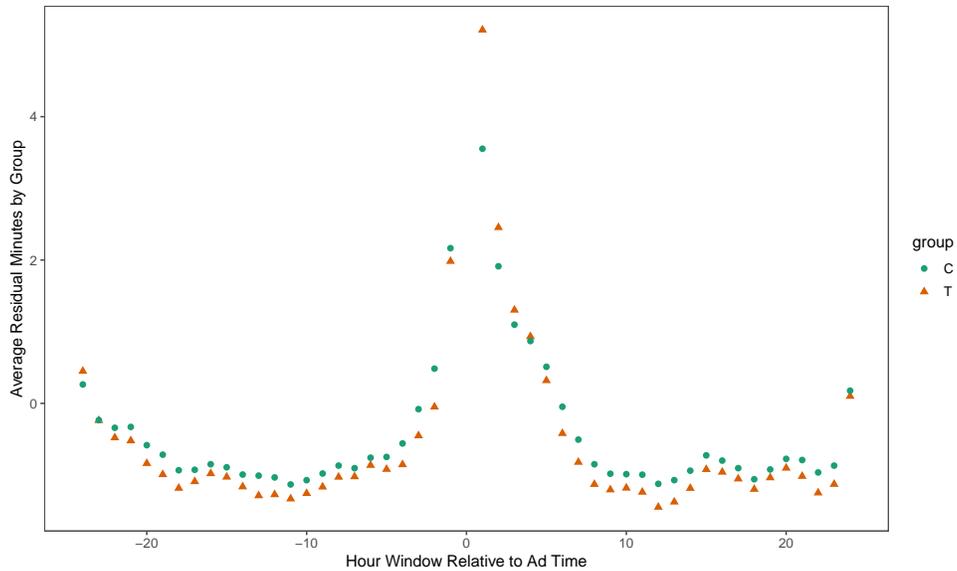
(B) Residual news viewing from saturated regression

Notes: We present two checks on the existence of pre-trends. In the first, we present the average news viewing across hourly bins for treatment and control groups. In the second, we present residual hourly news viewing from a regression that has fixed effects at the time-zone by hour-of-day levels. As such, it controls for time the ad aired and location heterogeneous effects. Additional details are presented in Appendix B.1

The observed temporal pattern visible in the figures provides evidence for cyclicality on this time scale.

Finally, we emphasize that anticipation of treatment due to unobservables that could explain the outcome (news viewing) is a priori implausible in our setting, given that "treatment" is defined as being tuned in at the instant an ad begins. Anticipation would require knowledge of the schedule of an ad block, which is unlikely.

3.2. Estimation

We estimate equation (3.1) by computing the Ordinary Least Squares estimator for δ_a at the ad-level, exploring our quasi-random variation. For a certain ad a, we compute $\hat{\delta}_a = \frac{1}{N_{a,T}} \sum_{i \,:\, T_{i,a}=1} (y_{a,i,1} - y_{a,i,0}) - \frac{1}{N_{a,C}} \sum_{i \,:\, T_{a,i}=0} (y_{a,i,1} - y_{a,i,0})$. This is just the average change in news viewing time for the treated group minus the average change in news viewing time for the control group for a given ad.

We exploit the fact that our research design is valid within each ad, and so equation (3.1) can be run at the ad-level. This also means we can estimate different $\{\delta_a\}_{a=1}^A$, across all ads $a = 1, ..., A$, identifying differential effects depending on ad characteristics. For example, we can classify ads within subgroups, such as by day of the campaign, time of the day in which it is aired, or the type of campaign it relates to. We can then aggregate all the $\{\delta_a\}_{a=1}^A$ within these particular subgroups to illustrate average weighted and equally weighted effects of campaign ads on news viewership across samples. For standard errors, we proceed by an appropriate bootstrap. We stack all estimates of δ_a together. We then bootstrap from this stacked dataset of estimates and recompute our estimates each time. This bootstrap is valid under the assumption of joint asymptotic normality. Additional details on this procedure and its statistical validity are available in Appendix B.2.

The procedure above presents multiple advantages relative to standard DiD or event-studies models in our set-up, as we cannot run model (3.1) jointly across ads to simultaneously estimate all $\{\delta_a\}_{a=1}^A$. In our case, we have tens of millions of observations, making computation of the estimators computationally difficult. Second, doing so would include some households changing treatment status across experiments. This is because treatment and control groups may overlap across ads: someone who watched an ad today might have not watched the

equivalent ad the next day.[26] Finally, since we can estimate our parameter of interest δ_a *for each ad* a, we do not suffer the same aggregation problems as dynamic DiD/event-study models have when trying to estimate a single model with multiple treatment effects.[27]

4. Results

Our results across 3 main specifications are shown in Table 2. We present average results across all ads in the first row under different weighting schemes. The other rows present heterogeneous effects by ad-type (whether the ad was sponsored by a candidate for a House race, a presidential one and so forth, and whether the candidate was an incumbent or a challenger). In the first 2 columns, we weigh ads according to treatment and control size - the first, as a geometric mean $\sqrt{n_{a,C}n_{a,T}}/\sum_a \sqrt{n_{a,C}n_{a,T}}$, the second by the number of devices $(n_{a,C} + n_{a,T})/\sum_a(n_{a,C} + n_{a,T})$. These are our preferred specifications, as they account for the sample size within each ad. The third column presents the results by averaging ads with equal weighting. Standard errors are block bootstrapped at the ad level, as explained in Appendix B.2.

Our results show that, on average, the treatment effect of a political ad is of approximately 3-4 minutes of news viewing the following day. This result is statistically significant and stable across specifications. As a benchmark, the median ad has average viewership of news among treated viewers in the previous 24h of approximately 74m, for an increase of about 5.4%.[28]

---

[26]Unfortunately, we are not aware of works in DiD with multiple time periods that can accommodate such a set-up. For example, Callaway and Sant'Anna (2020) assumes that once a subject is treated, (s)he remains treated throughout the sample, an assumption also maintained in Athey and Imbens (2018). This is clearly inappropriate in our set-up, where viewers in the treated group are not necessarily viewing the next ad. Meanwhile, the procedure that would most closely relate to ours is that of de Chaisemartin and D'HaultfŒuille (2017). Unfortunately, an application of their estimator for fuzzy DiD (i.e. agents can change group status) would require us to track the treatment status of every individual at every ad, which is computationally infeasible.

[27]For example, the latter requires appropriately defining the parameter of interest (the treatment effect relative to which control subgroup) and does not provide appropriate weighting of treatment groups under heterogeneous effects, see Goodman-Bacon (2018); de Chaisemartin and d'Haultfoeuille (2020) for instance. These are issues that do not come up in our framework.

[28]Providing a benchmark in this set-up is not straightforward, as there are many possible choices. The method here first averages previous-24h news viewing over the treatment group for a given ad, then takes quantiles of the resulting ad-level distribution. We view this as the closest analogue to our ad-level estimates. An alternative is to aggregate first within device (averaging across all ads for which the device appears in treatment) and

TABLE 2. Average Effects and Bootstrapped Confidence Intervals, by Subgroup.

	Subgroup	Effect (Minutes)		
		Inv. Variance	Total Devices	Equal Weight
		(1)	(2)	(3)
Full Sample		4.038	3.860	2.595
		(3.859, 4.208)	(3.689, 4.026)	(2.268, 2.913)
Sponsor	House	3.079	2.977	0.399
		(2.752, 3.403)	(2.647, 3.294)	(−0.206, 0.989)
	President	4.074	3.847	4.017
		(3.773, 4.404)	(3.551, 4.176)	(3.411, 4.633)
	Senate	4.572	4.312	2.932
		(4.082, 5.061)	(3.853, 4.779)	(2.058, 3.819)
	Statewide	6.824	6.784	3.659
		(6.316, 7.316)	(6.276, 7.277)	(2.839, 4.470)
	Outside Group	2.523	2.344	2.055
		(2.110, 2.946)	(1.950, 2.760)	(1.420, 2.797)
Incumbency	Challenger	5.067	4.918	3.516
		(4.758, 5.370)	(4.621, 5.210)	(2.890, 4.091)
	Incumbent	3.871	3.691	2.197
		(3.629, 4.117)	(3.453, 3.932)	(1.753, 2.661)

Notes: The table reports average effects, weighted by 1) the geometric mean of treatment and control group size, 2) total devices in treatment and control groups and 3) a simple equally weighted average. Confidence intervals are the central 95% interval of 1000 bootstrap replicates within each subgroup.

This result speaks to work such as Huber and Arceneaux (2007), which concludes that political advertisements do not engage voters into politics. Even though campaigns do not appear to affect turnout (as found in Ashworth and Clinton, 2007; Krasno and Green, 2008), a first measure of engagement, we find that viewers do increasingly "engage" with campaigns after ads through increased news viewing. Our mechanism is consistent with both the literature's results on lack of effects on turnout, as well as their evidence that advertisements are able to persuade voters. In particular, beliefs can change because there is an increased viewership of political news through campaigns, and this effect on media consumption responds to the type

---

then taking quantiles of the device-level distribution. In this method, the median treated device's previous-24h viewing is 14 minutes. The level of news viewing in our sample is higher than that of the average TV viewing population because, as noted previously, a substantial fraction of political ads run on news programs and hence our construction of treatment and control groups selects for types with above-average preference for news programs. Additionally, the period we study (less than 2 months prior to a presidential election) has elevated overall viewership of news programs.

and importance of those campaigns. These effects are significant, but possibly small enough that single ads might not generate observable effects on outcomes such as turnout. However, the combination of multiple ads over the campaign might. We explore the role of dynamics and dynamic effects from Section 4.2 onwards.

The results in Table 2 show evidence of some heterogeneity in treatment effects depending on characteristics of the sponsor. The estimates are larger in presidential, Senate and especially statewide elections, as well as for ads sponsored by challengers. For example, while the average effect for an ad sponsored by a candidate for the House is around 3 minutes in our preferred specifications, it is over 4 for Senate campaigns, and closer to 7 minutes for statewide campaigns. We also examine heterogeneity in characteristics of the audience. Table 3 shows results of OLS regressions of the ad-level estimated treatment effect $\hat{\delta}_a$ on average characteristics of the viewers in the ad's treatment and control groups.[29] Columns again show the different weighting schemes corresponding to those in Table 2. We compute ad-level average values of several demographic attributes, and then construct dummies for quartiles of the distribution of ad-level averages for each characteristic.[30] We additionally include fixed effects for the sponsor characteristics examined in Table 2.

The results show that estimated effects are lower, sometimes substantially so, for ads whose audience is at the high end of the distribution for age, income, and education. These are precisely the characteristics most associated with high levels of political information, implying that ads' stimulating effect on news consumption is most effective for the types of viewers who otherwise would consume relatively little political information.

Another dimension of heterogeneity we can explore is to compare effects across different news channels. We decompose our total news viewing outcome $y_{i,a,w}$ into channel specific news viewing outcomes $y_{i,a,w}^c$, where c indexes channels.[31] We recall that the source of ads in

---

[29]For purposes of computing average characteristics, we pool both groups together.

[30]The existence of ads with a handful of viewers in our tracking data means there are many extreme outliers of the ad-level average values, making a linear specification inappropriate.

[31]We break y into separate categories for each of the four main cable news channels CNN, Fox News Channel (FNC), Headline News (HLN), MSNBC; a combined category for the public channels CSPAN and CSPAN2; a combined category for any local affiliate of the broadcast networks; and a combined category for all other cable channels.

TABLE 3. Heterogeneity in DiD Estimates by Audience Characteristics

| | DiD Estimate | | |
| | Inv. Variance | Total Devices | Equal Weight |
	(1)	(2)	(3)
Age (Q2)	-2.172**	-2.131**	-0.3423
	(1.073)	(1.058)	(1.074)
Age (Q3)	-1.563	-1.551	-1.895
	(1.198)	(1.143)	(1.397)
Age (Q4)	-7.155***	-7.025***	-6.174***
	(1.354)	(1.313)	(1.178)
Income (Q2)	-0.3075	-0.4129	0.1422
	(0.2940)	(0.2900)	(0.6323)
Income (Q3)	0.6972*	0.5302	-1.007
	(0.4073)	(0.4605)	(0.6125)
Income (Q4)	-2.824***	-2.972***	-4.109***
	(0.1507)	(0.1912)	(0.7124)
College Grad (Q2)	1.089***	1.194***	0.8244*
	(0.3122)	(0.3244)	(0.4465)
College Grad (Q3)	-1.981***	-1.849***	-0.2802
	(0.4526)	(0.4786)	(0.7839)
College Grad (Q4)	-3.438***	-3.227***	-2.123*
	(0.8352)	(0.8873)	(1.117)
Black (Q3)	0.1966	0.2259	2.082**
	(0.5063)	(0.5111)	(0.9891)
Black (Q4)	-5.075***	-4.951***	-5.1***
	(1.18)	(1.173)	(1.805)
Hispanic (Q3)	-0.6127	-0.5896	-0.8096
	(0.4815)	(0.4706)	(0.8661)
Hispanic (Q4)	-1.624***	-1.68***	-3.149***
	(0.6246)	(0.6246)	(0.6095)
Office FE (5)	✓	✓	✓
Incumbent FE (2)	✓	✓	✓
Observations	78,526	78,526	78,526
R^2	0.005	0.005	0.006

Notes: Significance Codes: ***: 0.01, **: 0.05, *: 0.1. One-way (Office) clustered standard-errors in parentheses. An observation is an ad. The dependent variable is the ad-level differences-in-differences estimate of the effect on news consumption. The sample is all devices active and tuned in to the channel on which a political ad ran at the time the ad began, i.e. the treatment group from the differences-in-differences analyses. Column (1) weights by the geometric mean of treatment and control group size; column (2) weights by the total number of devices (treatment plus control group size); column (3) is equally weighted. Each audience characteristic is computed as the mean of the variable among all devices in treatment and control groups for a given ad. Regressors are dummies for the ad being in the indicated quartile of the distribution of ad-level means for the indicated characteristic. The first and second quartiles of the ad-level mean for Black and Hispanic are equal (the median is zero for both), hence the Q2 dummy is omitted for these variables.

our dataset are local affiliates, as there are no ads on national cable channels like Fox News in this dataset.[32] The results are presented in Table 4. They show that increases in the viewing of political news are predominantly in local news through local affiliates.

One explanation consistent with these results is that treated viewers have preferences for network affiliates to begin with. This is because treatment was defined exactly on those watching the local affiliates on which political ads ran. Hence, viewers seem to increase consumption of news on their initially-preferred channels. Table 4 also suggests that the increase in local news viewing is offset by a small amount of substitution away from national channels.

TABLE 4. Average Effects and Bootstrapped Confidence Intervals, by Channel.

| | | | | Effect (Minutes) | | | |
| | CNN | CSPAN | FNC | HLN | MSNBC | OTHER | NETWORK |
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treated	−0.411	−0.047	−1.153	−0.207	−0.369	−0.131	6.356
	(−0.617, −0.205)	(−0.253, 0.159)	(−1.359, −0.947)	(−0.413, −0.001)	(−0.575, −0.163)	(−0.337, 0.075)	(6.150, 6.562)

Notes: The table reports average effects of news viewership on the indicated channel, weighted by the geometric mean of treatment and control group size. OTHER is all other national cable networks, and NETWORK is all local network affiliates. Confidence intervals are the central 95% interval of 1000 bootstrap replicates within each subgroup.

One may be interested in the sources of these effects: are treated viewers increasing their news consumption because they are stimulated to actively search for new sources of information, or rather, the ad incentivizes them to spend further time on a news program which they are already watching? When we decompose our main specification results, we find an estimate of 3.12 minutes in political ads aired on non-news programs (about 58% of the data), smaller than the 4.97 minutes effect when conditioning on ads aired on news programs. However, the latter group's sample of viewers has a larger baseline of news viewing. Hence, there is evidence that ads induce active seeking of news on other programs, and not simply continuing to watch the same news program on which an ad aired.

By increasing news viewing over the campaign, the treated group also changes their likelihood of future treatment (exposure to more ads), an effect we explore in Section 4.2.

---

[32]Our treatment and control groups were defined on viewers watching ads on local TV at the time, so they would not be watching national channels concurrently.

4.1. Interpretation of Baseline Results within a Theoretical Framework

Our main result contrasts with many standard models of the effect of information on voter's beliefs. The standard Bayesian framework (e.g. Achen, 1992; Gerber and Green, 1998) would suggest that, with additional ads voters would want to decrease exposure to costly information. Typically in such models, ads would be modelled as (noisy) informative signals that are identically, normally and independently distributed. As the variance of viewers' beliefs tighten with each additional ad, the marginal benefit decreases, so more information might no longer be worth the cost of search. In contrast, we find a positive average effect — political ads are *stimulating* news viewership, and this effect is statistically significant. We can decompose our main estimates into hourly windows, following the method detailed in Section B.1. Figure 4 shows that the bulk of the effect happens shortly after the ad and decreases with time, consistent with the stimulation channel.

FIGURE 4. Decomposing the 24 hour difference-in-difference effect hour-by-hour.
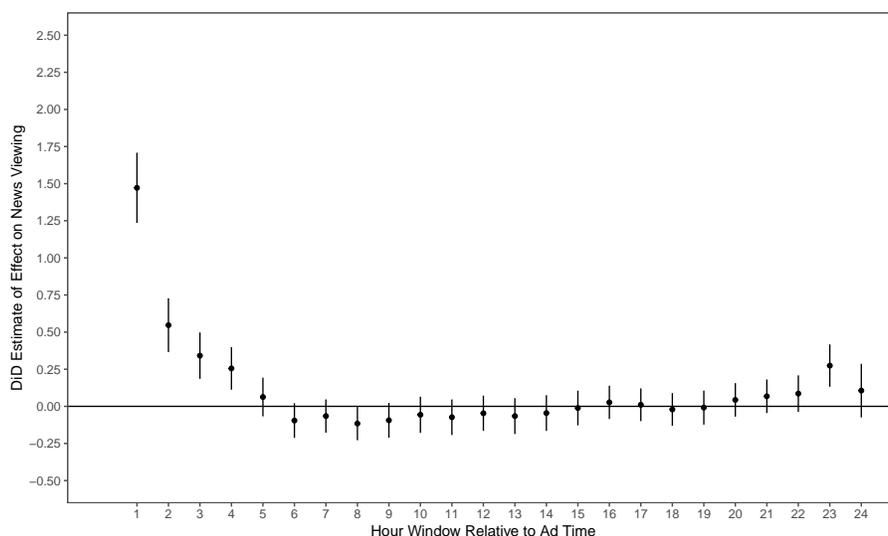
Notes: The figure decomposes our main effects into hourly effects. Dots are point estimates, vertical bars are 95% confidence intervals.

One explanation for our main results could be voter inattention, since unexpected ads appear to stimulate interest in news. However, the patterns of response indicate a sophisticated rather than naïve inattention. From the heterogeneity effects in Table 2, it appears that political ads have stronger effects for more salient campaigns with a higher marginal impact on

policy (gubernatorial, senatorial, presidential), and those of which voters are less likely to be aware of ex-ante (challengers).

Matějka and Tabellini (2020) provide a straightforward theoretical framework that can accommodate this variation. Their model of electoral competition with rationally inattentive voters has variation in the quality of prior information that voters possess as well as variation in the cost of acquiring new information across voters. An interpretation of our results consistent with this model is that advertisements reduce the acquisition-cost parameters for exposed voters. While the standard Bayesian model proposes that each voter has a fixed cost of acquiring information, in the rational inattention framework advertisements allow the politician to reduce a citizen's search costs. In our context, the most natural interpretation is that search costs take the form of cognitive or processing effort (McGraw, 2000). Ads can alter cognitive or processing demands of acquiring information by e.g. making viewers aware of new candidates and policies (an aspect tested in DellaVigna and Kaplan, 2007), increasing the perceived salience of certain political issues (Le Pennec and Pons, 2019), or making viewers aware of the intensity of political competition or of the preferences of their peers (Bursztyn et al., 2020).

Sponsors with greater resources or greater intrinsic interest — like presidential or gubernatorial candidates — may be more effective at achieving this search cost reduction. Furthermore, for a given search cost reduction, voter response will be larger for candidates about whom priors are more dispersed, e.g. challengers, and for voters with more dispersed priors, such as those with less political information and less political news consumption. This is what we find in our heterogeneity analysis above.[33]

4.2. Dynamics

Our main estimates in Table 2 average over the entire campaign. Given our proposed mechanism (that ad exposure increases interest in information-gathering about the campaign) we

---

[33]An alternative framework is that of cheap talk with multiple receivers (e.g. the theory in Farrell and Gibbons, 1989; Goltsman and Pavlov, 2011, and experimental work in Battaglini and Makarov, 2014). In this set-up, politicians are better informed than the public, and choose signals through ads to inform the latter. In practice, however, real world complexity creates a wedge between this framework and what we can measure — empirically, our receivers do not necessarily receive the signals from the sender, and we face an election in which multiple politicians (senders) compete among themselves when sending signals.

expect heterogeneity of effects over time. As the campaign goes on, some viewers may become saturated with political information and thus the stimulating effects of campaign ads may decline. To investigate this hypothesis, we plot the daily average effects $\{\bar{\delta}_1, \bar{\delta}_2, ..., \bar{\delta}_T\}$, disaggregated from Table 2.[34] The daily average effects are shown in Figure 5, along with a fitted linear trend line which shows that the average short-term effects on news viewing are declining as the election approaches.

FIGURE 5. Treatment effects across the campaign
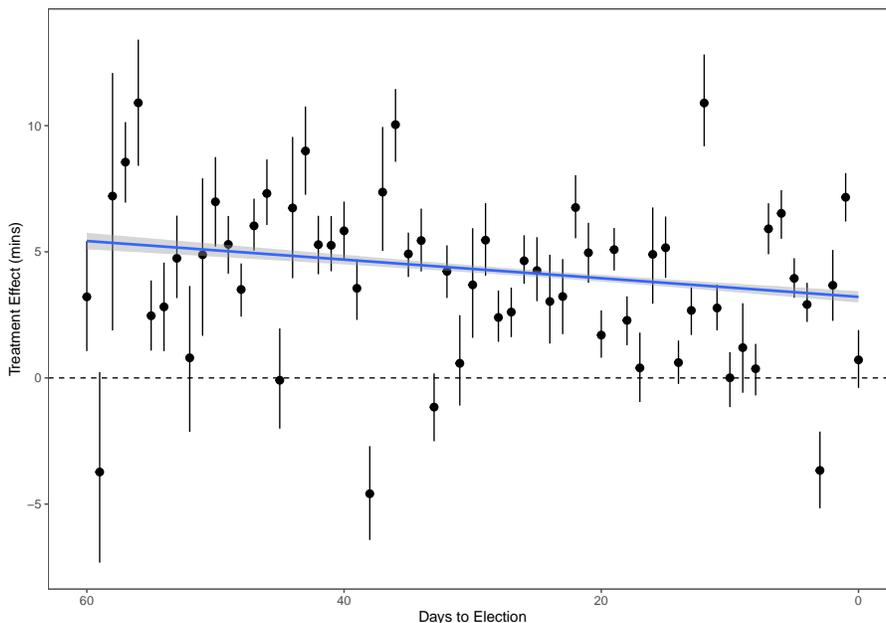
Notes: We present the average treatment effects at the daily level (average effects across all ads within a day, with each given an equal weight). 95% bootstrapped confidence intervals for ads by day. The blue line is the least squares fit of the estimated effect (at ad level) on days to election, weighted by the geometric mean of treatment and control group size. This regression has a statistically significant slope (coefficient= 0.03, SE = 0.006, t = 5.1), indicating that the treatment effects are largest with more days remaining in the campaign (i.e. left-hand side of the graph).

The short-term effect of an ad on news viewing in the next 24 hours in day t ($\bar{\delta}_t$) is decreasing as election day approaches.[35] However, the overall effect of an ad are decreasing

---

[34]The daily averages $\bar{\delta}_t$ are defined by $\bar{\delta}_t = \frac{1}{\sum_i w_{it}} \sum_i w_{it} \hat{\delta}_{it}$, where w_{it} is the geometric mean of treatment and control group sizes for ad i on day t.

[35]We note that this over-time comparison mixes together two possible changes: declining marginal effects of advertising within viewer (e.g. due to saturation) and an expanding population of exposed viewers as the volume and reach of advertising increases towards the end of the campaign. Figure B.4 in the Appendix shows

at an even faster rate. This is because in the beginning of the campaign, an additional 3-4 minutes of news in the next day increases the likelihood of exposure to additional political ads tomorrow and hence further stimulation of news consumption. Being exposed early on can therefore generate a chain of cumulative effects on news consumption; exposure at the end of the campaign produces only the direct (short-term) effect. Both of these effects point to the direction that early advertisement would be more effective than late ones, all else kept constant, motivating early spending.

To quantify this cumulative effect over the campaign, we estimate the same model as in equation (3.1) using the number of political ads a viewer is exposed to in a given time window (again, ± 24 hours around the initial ad air time) as the outcome. Hence, the treatment effect is how many more ads does one political ad lead treated viewers to watch relative to the control over the course of the next 24 hours. Results are shown in Table 5.

We find, as expected, a positive and significant effect on future ad exposure. A viewer in the treatment group will expect to be exposed to about 1-1.5 additional political ads, on average, compared to a viewer in control group, over the next 24 hours. As a result, the effect of early exposure is multiplied: news viewing also increases through exposure to additional future ads, which further increase news viewing.

TABLE 5. Average Effects and Bootstrapped Confidence Intervals, by Subgroup.

| | Effect (Additional Ads) | | |
| | Inv. Variance | Total Devices | Equally Weighted |
	(1)	(2)	(3)
Treated	1.267	1.325	1.727
	(1.079, 1.455)	(1.137, 1.513)	(1.539, 1.915)

Notes: The table reports average effects, weighted by 1) the geometric mean of treatment and control group size, 2) total devices in treatment and control groups and 3) a simple equally weighted average. Confidence intervals are the central 95% interval of 1000 bootstrap replicates within each subgroup.

---

that these two forces roughly balance each other out, preserving a similar fraction of not-yet exposed devices in the sample over the last month of the campaign. Since around 80% of devices in treatment or control groups in this period have been previously exposed, and such exposure is balanced across groups (as seen in Figure B.3), we conclude that the primary dynamic change comes from diminishing marginal effects.

To what extent can this mechanism explain the estimates for the effects of political advertising on voting outcomes found in the literature (e.g. Ashworth and Clinton, 2007; Krasno and Green, 2008; Spenkuch and Toniatti, 2018)? We cannot directly answer this question, as we have no individual voting records and cannot compare vote decisions of just-exposed versus non-exposed voters. Nevertheless, we can provide a very coarse back-of-the-envelope estimate of the possible magnitude. To do so, we can multiply our (average) estimate of four minutes additional news viewing over the 10 campaign weeks in our sample by the Martin and Yurukoglu (2017) estimated persuasive effect of viewing the Fox News Channel (FNC), of 0.15 p.p. increase in Republican voting probability per minute per week of FNC viewed. This calculation approximates the effect for a candidate favored by news coverage as much as Republican presidential candidates are favored by FNC — e.g., very likely an upper bound for our setting. The standard deviation of the total number of ads a viewer is exposed to in our data is about 30,[36] which implies that a 1 standard deviation increase in ad exposure could lead to a 1.8pp increase in vote share through increased news consumption, for a candidiate with favorable news coverage. This estimate is larger than the Spenkuch and Toniatti (2018) estimate of 0.5pp per 1 s.d. increase.[37] While this is a rough upper bound, at the very least it suggests that this mechanism could be generating measurable effects on voting outcomes.

4.3. Placebo Tests and Robustness Checks

We probe the validity of the main results in the difference-in-differences design through a series of placebo tests and robustness checks. These tables and figures can be found in Appendices C-D.

---

[36]The distribution is very skewed, with the median household being exposed to four political ads. Furthermore, this value is 50% larger than the 22 ads in the sample of Spenkuch and Toniatti (2018).

[37]There are at least two reasons to think this calculation using FNC persuasion rates is an upper bound for our case. First, as Table 4 shows, all of the increase in news consumption comes from network news programs, which are much less partisan than FNC. Second, the Martin and Yurukoglu (2017) estimate is the LATE on viewers induced into watching FNC by channel position, who are likely to be viewers with relatively weak preexisting partisan affiliation and who are hence likely to be relatively highly persuadable. We expect this effect to be larger than an average across a larger population — by comparison, our estimate is an ATT on all viewers of programs on which political ads air.

First, we validate our main estimates from Table 2 by checking whether those effects are present in non-news programs. To do so, we run our main specification (3.1) on (i) viewership of sports programs, and (ii) viewership of comedy programs (primarily sitcoms, but also including late-night variety shows) in the 24 hours following treatment. If the mechanism underlying our effect is indeed a stimulation of interest in the campaign, we should not expect to see an increase in viewing of non-news programs. The first row of Tables C.2 and C.3 in Appendix C.2 confirm this: there is no evidence that political ads increase viewership of either class of programs. The effect on consumption of sports is approximately zero (or slightly negative), and negative on comedy shows. This is consistent with viewers increasing their news consumption by substituting away from non-news channels/programs, as shown in Table 4. Political ads increase interest in news specifically, not TV viewing in general.

Second, we probe our main results by checking how they change with different definitions of the control group. Specifically, we decompose the control groups into three subgroups: those who tuned out before the ad came on, those who tuned in after the ad aired, and those who did both. While our design eliminates confounding fixed differences in news viewing levels across viewers as well as cyclical differences in viewing trends, in the absence of explicit randomization, it remains open to possible bias from short-term differences in pre-exposure trends between treated and control viewers. We expect that such bias likely has the opposite sign for viewers who tuned out prior to the ad air time (and thus were likely already reducing their news consumption — noted as group C1) versus viewers who tuned in after ad air time (and thus were likely increasing their news consumption, group C2). These estimates thus are expected to bracket the true effect (in a partial identification sense). We show the results in Table C.1.

We emphasize that we do not view this exercise as a sensitivity analysis where we expect, if the results are not spurious, for the differences between estimates under the different definitions to be zero. Rather, this split is inspired by the partial identification literature on treatment effects (e.g. Flores and Flores-Lagunes, 2013), and gives us bounds on the estimates that are robust to the main threat to identification. Indeed, we find that our main estimates from

Table 2 are within the bounds $[2.5, 4.6]$ suggested by Table C.1 and our main results are positive and statistically significant throughout the specifications, regardless of which control group is used.[38] In addition to its value in bounding the object of interest in our paper, this technique is transferable to other difference-in-differences settings with similarly high-frequency data.

Shrinking the window defining treatment from 20 minutes to 5 and 10 minutes still yields positive and significant estimates that are close to the baseline results, whether splitting control groups (Tables C.4-C.5) or not (Tables C.6-C.7). Appendix C.3 shows that our main results are robust to re-balancing the covariates Age and Hispanic, a test suggested by Figure B.2 and alluded to in Section 3. Finally, in Appendix D, we find that our main results actually increase when we drop ads with missing data, suggesting that such incomplete data is not biasing our estimates upwards.

4.4. Additional Support from a Regression Discontinuity Design

We now complement the difference-in-differences evidence by exploring a Regression Discontinuity Design (RDD). Our running variable is defined by the amount of seconds since an ad is aired: varying from $-\tau$ seconds up to τ, with treatment occurring at 0. As outcomes, we use both the amount of news viewing in the next 24 hours, as well as the change in news viewing over the next 24 hours relative to the previous 24 hours.

The RDD allows us to check whether our effect is still present under alternative identification assumptions and in an alternative (albeit related) parameter of interest, a local average treatment effect. Furthermore, it provides a formal way to estimate the model under a much tighter windows around the ad. On the other hand, to implement this design, we must pool all ads for statistical power, limiting our results on heterogeneity.

The RD design is valid under the well known assumptions of the absence of manipulation of the running variable, and the continuity of unobservables around the cutoff. In our case, these identification assumptions are likely to hold: as discussed in the previous section, TV viewers

---

[38]Using the same decomposition, we find that the placebo tests from Tables C.2-C.3 are also robust to the definition of the control group. The average effect is close to 0 or in the opposite direction to our estimates. They are also robust to using the control group most likely to invalidate our design (C2 — those tuning in after the ad and likely to be increasing their news consumption).

are unlikely to perfectly anticipate the entry of a political ad. This argument is even stronger when the bandwidth is only a few seconds. Furthermore, we have shown that treatment and control groups are balanced along covariates in the previous section, a common check in regression discontinuity designs. Together with the high frequency dataset, this suggests that it is improbable that there is selection on unobservables.

We first present the specification in visual form. Figure 6a plots 1-second binned averages of news viewing time in the 24 hours following the ad against the running variable, tune-in time relative to ad air time — e.g. the point at -50 is the average news viewing time in the 24h subsequent to ad air time, among all devices which tuned in exactly 50 seconds before the ad aired. Here we pool data across all ads to construct these binned averages. Devices on the negative side of the axis are the treated (T) group from the previous sections: they tuned in prior to the ad and were watching at ad time. Devices on the positive side are the C2 group: they tuned in after the ad start time.

We see a notable drop at time zero, indicating that the devices which tuned in just prior watch *more* news in the subsequent 24h than those who tuned in just after.[39] Note that there is a decline in average viewing time as we approach zero from the left; this is because being in the T group conditions on still watching at time zero, and hence the group that tuned in at -600s necessarily watches a longer stretch of the surrounding program than the group that tuned in at -60s. Viewing extended segments of the same program is likely correlated with greater overall time spent watching. The same trend is not evident on the positive side of zero; this is because there is no similar condition imposed on the segment length viewed by the C2 group.

Figure 6b plots the same binned averages computed over tune-in times relative to ad air-time, but uses news viewing in the *previous* 24h as the outcome variable.[40] Here we see the

---

[39]Note here that the regression discontinuity is not perfectly sharp, as those who tuned in seconds before may not have been paying enough attention to absorb the ad, and those who tuned in seconds after still see part of the ad. For this reason we see a steep but not vertical decline from the fully treated ($< 0s$) to fully untreated ($> 30s$) viewers.

[40]This placebo check is related to the "case-crossover" method used in epidemiology. In that empirical design, one uses a subject as their own control: their past observations at comparable times are a control for their outcomes after treatment, see Maclure and Mittleman (2000) for a survey. For example, Redelmeier and Tibshirani (1997) studies the effect of cell phone use on car crashes by comparing a subject who received a call when driving, to

same relationship between tune-in time and news viewing among treated and control (C2) viewers, but no discontinuity at zero. This combination of results provides evidence that the pre-to-post change we observe among the T group relative to the C2 group is not a spurious consequence of our design selecting viewers whose news viewing was already rising, but rather is actually a consequence of the ad exposure. We note that this placebo check could not be done in the difference-in-differences specification because the latter already differenced out past news viewing within each group. Finally, Figure 7a plots the difference of Figures 6a and 6b, i.e. the difference between the viewer's news viewing in the 24 hours after ad time and the same viewer's news viewing in the 24 hours prior to ad time. We see a notable discontinuity at zero, confirmed by the statistically significant estimates of around 2 minutes using the robust bias-corrected estimator of Calonico et al. (2014) across bandwidth choices, presented in Table C.10 in Appendix.

The RDD permits an additional validity check: we can compare our C1 group (viewers who tuned out before the ad aired) to the C2 group (viewers who tuned in afterwards). Neither of these groups were exposed to the ad, but they tuned in to the ad-containing program at similar times to each other. They thus provide a useful comparison to rule out that our results are driven by spurious differential time trends related to tune-in time. This is shown in Figure 7b which replicates Figure 7a but replacing the T group by C1. There is, reassuringly, no discontinuous drop in the pre-to-post viewing change in this comparison.

Altogether, the regression discontinuity design confirms and further validates our findings of campaign ads being able to spur viewers into consuming further political news.

5. Tune-Out

Our results to this point are all in intent-to-treat (ITT) terms: we define "treatment" as having been tuned in to the program on which the ad aired at the instant the ad *began*. It is possible for viewers to exhibit "noncompliance" by tuning out as soon as they realize a political

---

their own behavior at a comparable time in the previous day (when they were driving but no call was received). Our exercise differs, at least relative to the simplest implementation of this design, by the use of additional control groups (C2), as well as by the very high frequency and small bandwidth of our variation.

FIGURE 6. Visual regression discontinuity plots of news viewing against the running variable (tune-in time): Post-ad vs. Pre-ad viewing.
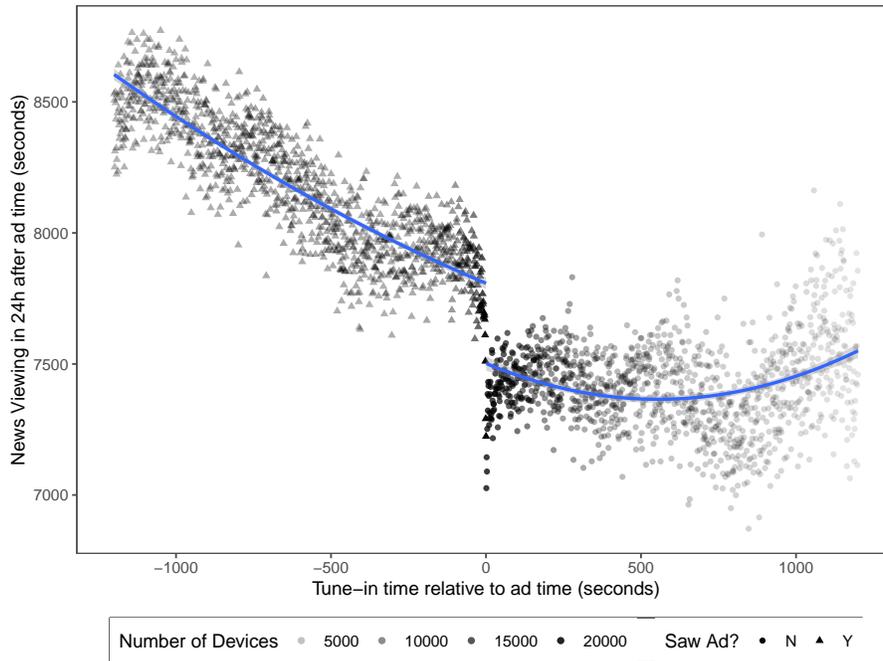
(A) Next 24h.
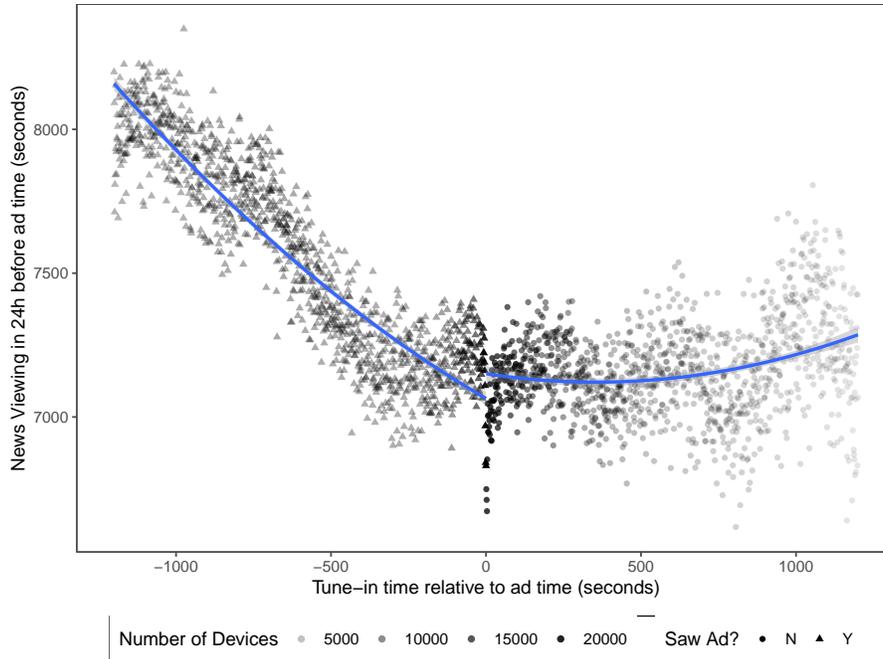
(B) Previous 24h.

Notes: Binned averages of news viewing time in the 24h *following* ad air time (A) versus the 24h *preceding* ad air time (B). Each dot is the average value for viewers in either Treated (T) or Control (C2) groups that tuned in x seconds after ad time (with negative values indicating tune-in prior to airtime and hence exposure to the ad.) The lines are quadratic fits, estimated separately on either side of zero.

FIGURE 7. Visual regression discontinuity plots of change in news viewing against the running variable (tune-in time): treated vs. control groups.
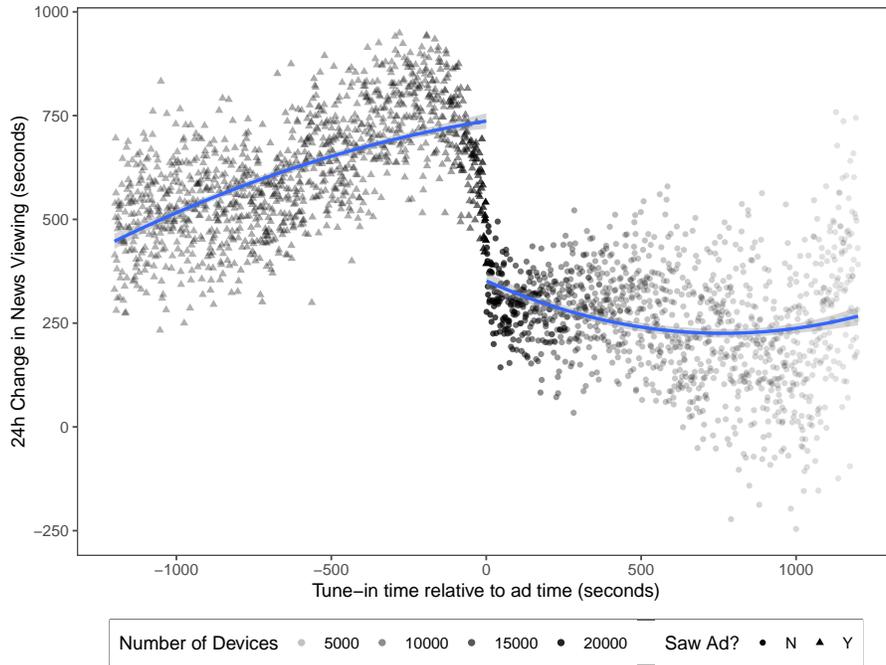
(A) T vs. C2.
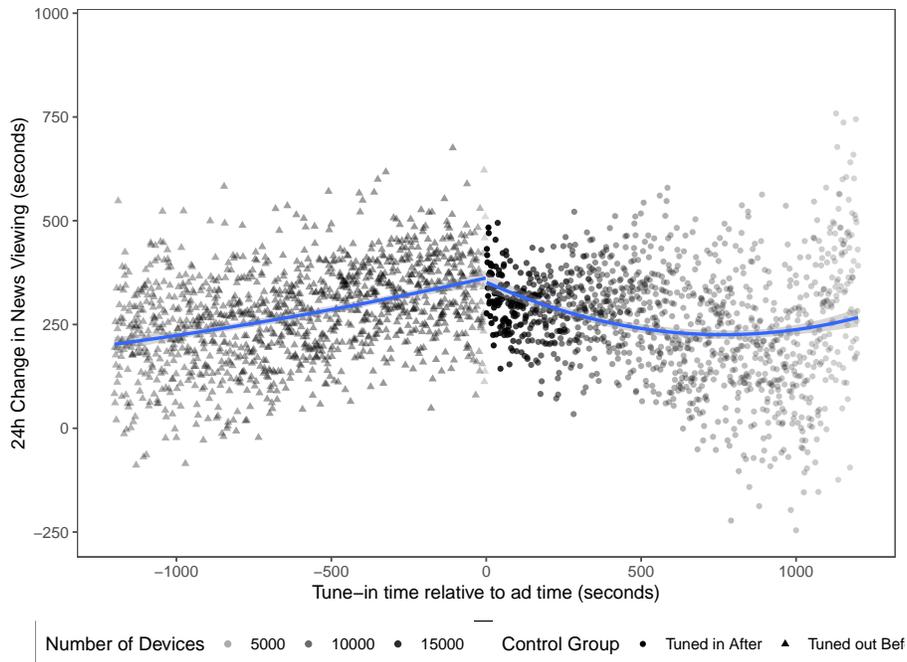
(B) C1 vs. C2.

Notes: Binned averages of *change in* news viewing time in the 24h following ad air time compared to the 24h *preceding* ad air time. Each dot is the average value for viewers in either Treated (T) or Control (C2) groups (panel A) or either C1 or C2 groups (panel B) that tuned in x seconds after ad time (with negative values indicating tune-in prior to airtime and hence exposure to the ad among T viewers, but not C1 viewers, who tune out again before ad airtime). The lines are quadratic fits, estimated separately on either side of zero.

ad has come on. In this section, we investigate viewers' choices to select out of exposure to political ads, by focusing attention on the set of treated devices and examining (variation in) the likelihood of sitting through the full ad without changing the channel, conditional on having been viewing at the instant the ad began. In other words, this allows us to check whether most treated viewers watch the full ad, and whether our ITT estimates would be close to an average treatment effect on the treated (ATT) (in the difference-in-differences design). We can also check if there is heterogeneity in the amount of treatment exposure across viewer characteristics. This analysis illuminates another source of dynamic variation in the strategic incentives to advertise over the course of the campaign.

Overall, we find that tune-out rates for political ads are quite low. The mean tune-out rate for political ads (defined as the fraction of viewers tuned in at the start of an ad who tune to a different channel before the ad ends) is 4.2%.[41] Appendix E shows that this rate is very similar to a synthetic comparison tune-out rate we construct using other times during the same day when the devices in the treatment group for a given ad were active. In this comparison set of non-ad times, the average tune-out rate is 4.6%. Viewers thus avoid political ads at rates that are low in absolute terms and similar to the average content that the treated groups consume at other times. See Appendix E for details.

Having shown that the average level of "noncompliance" is low, we move to examining heterogeneity. We first note that tune-out rates at the ad-level are negatively and significantly correlated with the ad-level effect size.[42] This provides some further corroboration that the effect is actually induced by exposure to the ad and not spurious (with the caveat that this relationship may reflect compositional differences in the audience rather than differences in the quality of ad content).

---

[41]By comparison, Knight and Tribin (2018) study tune-out in Venezuela, and find tune-out rates from Chavez's interventions on broadcasts to be an order of magnitude above ours. This is understandable, given that those interventions could be hours long (speeches by Chavez) and occurred in a much less politically competitive context.

[42]We regress the ad-level effect from the DiD analysis on the ad-level tune-out rate along with date fixed effects. The estimated magnitude is such that a 1-SD increase in tune-out rate (equal to about 9 percentage points) is associated with about a 2-minute drop in the estimated effect on news consumption. This relationship is highly significant with a t-statistic greater than 7 (using standard errors clustered by date).

Next, Table 6 shows results from regressions where the outcome is an indicator for a viewer tuning out before the end of a political ad, conditional on being active and tuned in to the ad's channel at air time (i.e., conditional on being in the "Treated" group for that ad). The columns split the sample by household partisanship: column (1) is households with unknown or independent partisanship; column (2) is households with a Republican-registering head of household, and column (3) is households with a Democratic-registering head of household.

TABLE 6. Differential Tune-out by Sponsor Characteristics and Timing.

| | Tuned Out | | |
	(1)	(2)	(3)
Dem Sponsor	0.0015***	0.0003	0.0002
	(0.0004)	(0.0009)	(0.0005)
Rep Sponsor	-0.0013***	-0.0031***	-0.0008
	(0.0003)	(0.0009)	(0.0005)
Days to Election	-0.0002***	-0.0002***	-0.0002***
	(7.4×10^{-6})	(1.784×10^{-5})	(1.068×10^{-5})
Subsample	Indep / Unknown	Rep	Dem
Household FE	✓	✓	✓
Observations	4,495,846	692,868	1,984,779
# Households	87,473	21,178	45,247
R^2	0.07301	0.08347	0.07607

Notes: Significance Codes: ***: 0.01, **: 0.05, *: 0.1. Standard errors (clustered by household) in parentheses. An observation is a household-ad. The sample is all households active and tuned in to the channel on which a political ad ran at the time the ad began, i.e. the treatment group from the differences-in-differences analyses. Column (1) restricts to households with independent or unknown party affiliation; column (2) restricts to Republican-identifying households, and column (3) restricts to Democratic-identifying households. The omitted category is outside group ads.

Across all three groups, the time trend is negative and statistically significant, indicating that tune-out is more likely the fewer days there are remaining until election. These regressions all include household fixed effects, indicating that the same viewer becomes less likely to sit through a political ad as the election approaches. This is consistent with increasing saturation with campaign information as the election approaches, and, in combination with our results on the decline in ad effects on news consumption over time (Figure 5) add an additional dynamic component to the incentives to advertise.

The first two rows in Table 6 show the effect of sponsor party on tune-out probability for each partisan group. There is some evidence of selective exposure on the partisan dimension. This selective response is asymmetric: Republican viewers are significantly less likely to tune-out of a Republican candidate's ad, whereas Democrats do not differentially select into exposure to Democratic candidates' ads. This asymmetry aligns with experimental findings on selective exposure in Henderson and Theodoridis (2018). Independents and viewers of unknown partisanship look more similar to Republicans in this sample (likely reflecting the baseline rightward skew of the markets included in the dataset, which tend to encompass smaller regional cities and surrounding suburbs and rural areas).

Overall, though, the magnitude of heterogeneity in selective exposure on both partisan and time dimensions is small. The party effects are comparable to moving an additional 5-15 days away from the election, and party explains only a small amount of variation in tune-out. Two ads on the same program, one sponsored by a Republican and one by a Democratic candidate, would have almost the same effective audience. And an ad 50 days out would see only about 1 percentage point less tune-out than one the day prior to election.

6. Conclusion

In this paper, we have provided evidence that exposure to political ads on television increases viewership of television news by approximately 3-4 minutes over the next day using quasi-random variation. Our effect appears robust across specifications, and does not appear to be driven by imbalance on predetermined characteristics or by differential pre-exposure viewing trends. With very fine data, we can look at the effects across different periods of the campaign and time of the day. The effect is strongest from ads in presidential and statewide races, stronger in ads sponsored by challengers, and is particularly pronounced among viewers who are on average less politically informed. Altogether, our results are consistent with viewer search costs for political information (or, equivalently, interest in the campaign) being altered by advertisements — a new mechanism. This suggests a role for advertisements

as changing the distribution of voters who are politically informed; not directly, but through inducing voters to inform themselves from other sources.

The dynamic variation in strategic incentives to advertise we uncover helps to explain the empirical fact that campaigns do not concentrate activity at the very end of a campaign, despite high-quality empirical evidence (Gerber et al., 2011; Kalla and Broockman, 2018) of rapid decay in ads' persuasive potential. There are strong trade-offs involved in waiting: campaigns miss the option to induce information acquisition at times when media signals are most favorable, and may find themselves at the end of the campaign facing an audience already saturated with campaign news and uninterested in hearing their message. The results that ads induce additional exposures to future ads, and that the news consumption-stimulating effect decreases over time, while tune-out rates increase, all push in the direction of advertising early.

The proposed channel allows for the effects of advertising advantages by a candidate to affect election outcomes (Gordon and Hartmann, 2013; Spenkuch and Toniatti, 2018) even in the absence of any direct persuasive effect of ads themselves. While ads might not be persuasive on their own, they stimulate information gathering; and the latter can be effective in persuading viewers. This mechanism could generalize beyond political economy, for example in rationalizing the mixed empirical evidence on the effectiveness of advertisements for consumer persuasion (DellaVigna and Gentzkow, 2010, p.649-650): product advertising has been the subject of discussions akin to the "minimal effects hypothesis" that motivated our work. Similar methods to ours could also be applied in that context, where similar high-frequency data is available.

Given the ebb and flow of political news throughout a campaign, the timing of advertising also has the potential to affect *what* voters learn. Our results suggest that candidates should advertise when news about them is good, and avoid it when news is bad. Even if campaigns cannot control the timing of information releases in the media (as in Gratton et al., 2017) they retain some control over which releases voters are likely to actually absorb. There is ample qualitative evidence that campaigns do engage in such coordination between the news cycle

and advertising messages. For example, Sides and Vavreck (2014) document that the Obama campaign launched a substantial wave of ads in July 2012 hitting Romney for his record at private equity firm Bain Capital following a July 12 Boston Globe story about Romney's time at the firm (pp. 124-125). To the extent that campaigns do have private information which they can choose to release to the media (such as announcements of endorsements, or a new policy proposal), our results suggest an incentive to coordinate the timing of such releases with advertising. As there is an incentive to release good news early (Gratton et al., 2017), there is also a corresponding incentive to advertise early.

On the methodological front, our work engages with the complexities of aggregating event studies with heterogeneous effects. With many quasi-experiments and a large panel, the implementation of improved event-studies estimators is far from trivial. As datasets increase, we expect that developing new methodologies might prove useful for such contexts.

While we do not provide a formal theoretical model, our results are informative for models of electoral competition and dynamic information acquisition. They also illustrate the complexity of fully embodying such campaign dynamics. A model of dynamic effects would have to incorporate multiple layers. First, politicians may anticipate that their ads will be watched by certain groups of viewers, and also anticipate that the distribution of such viewers will then change over time as a result. Second, there are multiple politicians who might be competing with one another when sending ads, not just over content or quantity but also on the dimension of congruence with media reports. Third, in contrast to existing models, politicians may target voters who are not actively acquiring information in a campaign, with the intention of engaging them. Finally, viewers might have different degrees of prior information and baseline interest in political information. While outside of the scope of this work, this research agenda seems promising to us.

References

Abraham, S. and Sun, L. (2020). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*.

Achen, C. H. (1992). Social psychology, demographic variables, and linear regression: Breaking the iron triangle in voting research. *Political behavior*, 14(3):195–211.

Ashworth, S. and Clinton, J. D. (2007). Does advertising exposure affect turnout? *Quarterly Journal of Political Science*, 2(1):27–41.

Athey, S. and Imbens, G. W. (2006). Identification and inference in nonlinear difference-in-differences models. *Econometrica*, 74(2):431–497.

Athey, S. and Imbens, G. W. (2018). Design-based analysis in difference-in-differences settings with staggered adoption. Technical report, National Bureau of Economic Research.

Avis, E., Ferraz, C., Finan, F., and Varjão, C. (2020). Money and politics: Estimating the effects of campaign spending limits on political entry and competition. Technical report, UC Berkeley, mimeo.

Battaglini, M. and Makarov, U. (2014). Cheap talk with multiple audiences: An experimental analysis. *Games and Economic Behavior*, 83:147–164.

Berelson, B. R., Lazarsfeld, P. F., McPhee, W. N., and McPhee, W. N. (1954). *Voting: A study of opinion formation in a presidential campaign*. University of Chicago Press.

Bursztyn, L., Egorov, G., and Fiorin, S. (2020). From extreme to mainstream: The erosion of social norms. *American Economic Review*, 110(11):3522–48.

Callaway, B. and Sant'Anna, P. H. (2020). Difference-in-differences with multiple time periods and an application on the minimum wage and employment. *Journal of Econometrics*.

Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6):2295–2326.

Chiang, C.-F. and Knight, B. (2011). Media bias and influence: Evidence from newspaper endorsements. *The Review of Economic Studies*, 78(3):795–820.

Coppock, A., Hill, S. J., and Vavreck, L. (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).

Cruz, C., Keefer, P., Labonne, J., and Trebbi, F. (2020). Making policies matter: Voter responses to campaign promises. Technical report, University of British Columbia, mimeo.

de Chaisemartin, C. and d'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–96.

de Chaisemartin, C. and D'HaultfŒuille, X. (2017). Fuzzy differences-in-differences. *The Review of Economic Studies*, 85(2):999–1028.

DellaVigna, S. and Gentzkow, M. (2010). Persuasion: empirical evidence. *Annu. Rev. Econ.*, 2(1):643–669.

DellaVigna, S. and Kaplan, E. (2007). The fox news effect: Media bias and voting. *The Quarterly Journal of Economics*, 122(3):1187–1234.

Farrell, J. and Gibbons, R. (1989). Cheap talk with two audiences. *The American Economic Review*, 79(5):1214–1223.

Ferraz, C. and Finan, F. (2008). Exposing corrupt politicians: the effects of brazil's publicly released audits on electoral outcomes. *The Quarterly Journal of Economics*, 123(2):703–745.

Flores, C. A. and Flores-Lagunes, A. (2013). Partial identification of local average treatment effects with an invalid instrument. *Journal of Business & Economic Statistics*, 31(4):534–545.

Fouirnaies, A. (2018). How do campaign spending limits affect electoral competition?: Evidence from great britain 1885-2010. Working Paper, University of Chicago, mimeo.

Gerber, A. and Green, D. P. (1998). Rational learning and partisan attitudes. *American Journal of Political Science*, 42:794–818.

Gerber, A. S., Gimpel, J. G., Green, D. P., and Shaw, D. R. (2011). How large and long-lasting are the persuasive effects of televised campaign ads? results from a randomized field experiment. *American Political Science Review*, 105(1):135–150.

Gerber, A. S., Karlan, D., and Bergan, D. (2009). Does the media matter? a field experiment measuring the effect of newspapers on voting behavior and political opinions. *American Economic Journal: Applied Economics*, 1(2):35–52.

Goltsman, M. and Pavlov, G. (2011). How to talk to multiple audiences. *Games and Economic Behavior*, 72(1):100–122.

Goodman-Bacon, A. (2018). Difference-in-differences with variation in treatment timing. Technical report, National Bureau of Economic Research.

Gordon, B. R. and Hartmann, W. R. (2013). Advertising effects in presidential elections. *Marketing Science*, 32(1):19–35.

Gratton, G., Holden, R., and Kolotilin, A. (2017). When to drop a bombshell. *The Review of Economic Studies*, 85(4):2139–2172.

Henderson, J. A. and Theodoridis, A. G. (2018). Seeing spots: Partisanship, negativity and the conditional receipt of campaign advertisements. *Political Behavior*, 40(4):965–987.

Huber, G. A. and Arceneaux, K. (2007). Identifying the persuasive effects of presidential advertising. *American Journal of Political Science*, 51(4):957–977.

Kalla, J. L. and Broockman, D. E. (2018). The minimal persuasive effects of campaign contact in general elections: Evidence from 49 field experiments. *American Political Science Review*, 112(1):148–166.

Kendall, C., Nannicini, T., and Trebbi, F. (2015). How do voters respond to information? evidence from a randomized campaign. *American Economic Review*, 105(1):322–53.

Knight, B. and Tribin, A. (2018). The limits of propaganda: Evidence from chavez's venezuela. *Journal of the European Economic Association*, 17(2):567–605.

Krasno, J. S. and Green, D. P. (2008). Do televised presidential ads increase voter turnout? evidence from a natural experiment. *The Journal of Politics*, 70(1):245–261.

Lazarsfeld, P. F., Berelson, B., and Gaudet, H. (1944). *The people's choice.* Duell, Sloan & Pearce.

Le Pennec, C. and Pons, V. (2019). Vote choice formation and the minimal effects of tv debates: Evidence from 61 elections in 9 oecd countries. *Harvard Business School Working Paper*.

Lovett, M. and Peress, M. (2015). Targeting Political Advertising on Television. *Quarterly Journal of Political Science*, 10(3):391–432.

Lovett, M. J. and Staelin, R. (2016). The role of paid, earned, and owned media in building entertainment brands: Reminding, informing, and enhancing enjoyment. *Marketing Science*, 35(1):142–157.

Maclure, M. and Mittleman, M. (2000). Should we use a case-crossover design? *Annual review of public health*, 21(1):193–221.

Martin, G. J. and Yurukoglu, A. (2017). Bias in cable news: Persuasion and polarization. *American Economic Review*, 107(9):2565–99.

Matějka, F. and Tabellini, G. (2020). Electoral Competition with Rationally Inattentive Voters. *Journal of the European Economic Association*.

McGraw, K. M. (2000). Contributions of the cognitive approach to political psychology. *Political Psychology*, 21(4):805–832.

Redelmeier, D. A. and Tibshirani, R. J. (1997). Association between cellular-telephone calls and motor vehicle collisions. *New England Journal of Medicine*, 336(7):453–458.

Ridout, T. N. and Smith, G. R. (2008). Free advertising: How the media amplify campaign messages. *Political Research Quarterly*, 61(4):598–608.

Sides, J. and Vavreck, L. (2014). *The Gamble: Choice and Chance in the 2012 Presidential Election*. Princeton University Press.

Spenkuch, J. L. and Toniatti, D. (2018). Political Advertising and Election Results. *The Quarterly Journal of Economics*, 133(4):1981–2036.

Vavreck, L. (2009). *The message matters: The economy and presidential campaigns*. Princeton University Press.

Appendix A. Additional data details

FIGURE A.1. Example of device-level data
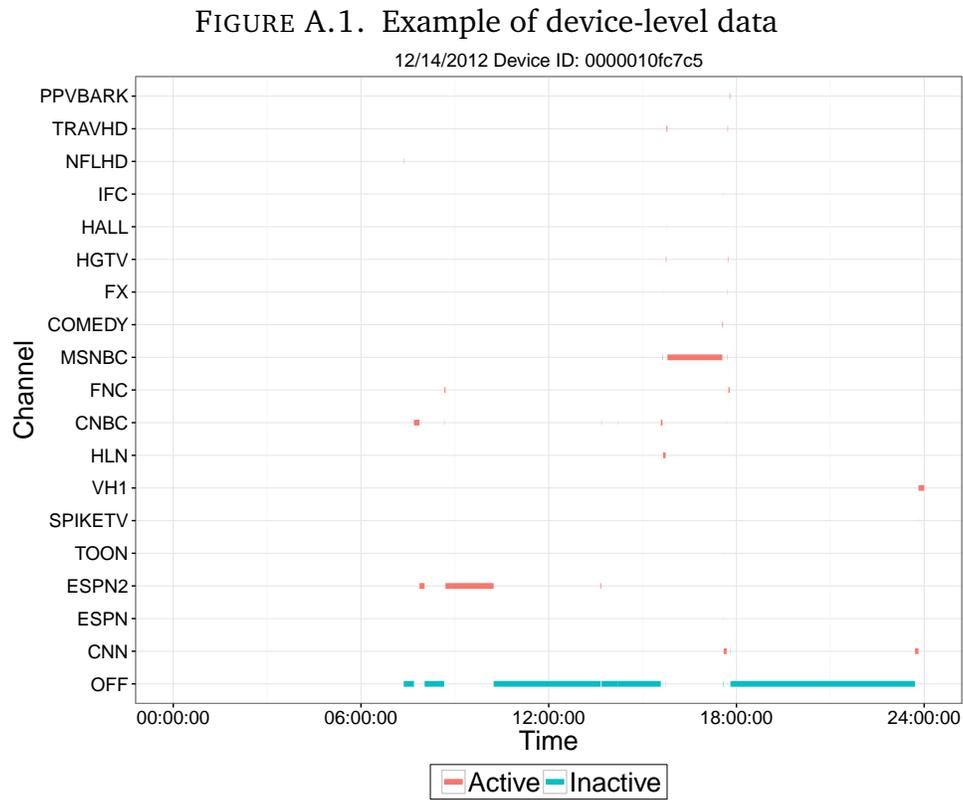
Notes: The figure presents an example of viewing data at the set-top box level for one device in our sample on one day, 12/14/2012. Active periods denote times in which the device is on. The vertical axis indicates the channel to which the device was tuned at a given time.

FIGURE A.2. Daily Spending on TV ads (in our data) during the campaign
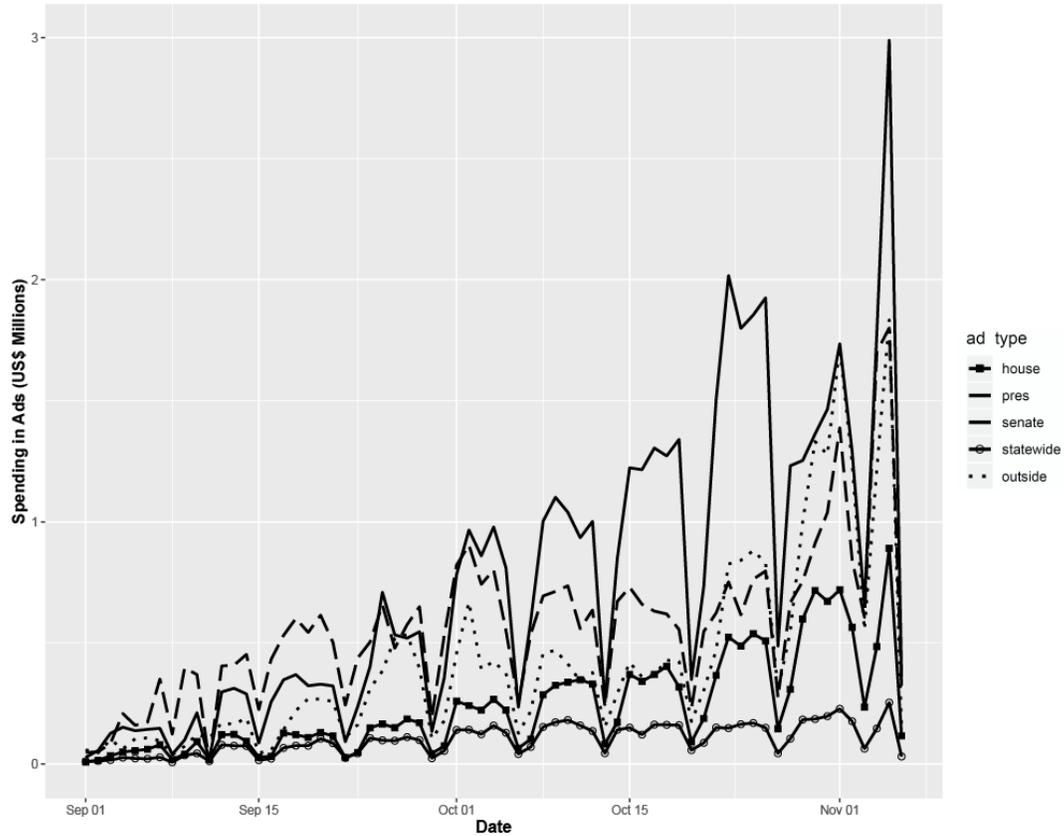
Notes: The figure presents the aggregate expenditure in ads in our data. Over 80% is spent on ads more than 2 weeks before election day. The cost of an ad is calculated by multiplying the cost per impression among adults 18 and over by the number of impressions for that ad. Each line is a category of sponsor: US House, President, Senate, State-wide (e.g. Governor) and outside groups.

TABLE A.1. Names of DMA's in our final sample, and in Figure 1

Name of DMA		
Abilene-Sweetwater	Alexandria, LA	Amarillo
Augusta	Austin	Bent, OR
Bluefield-Beckley-Oak Hill	Charleston - Huntington	Charlotte
Columbus, GA	Dallas - Ft. Worth	Detroit
Elmira	Eureka	Greenwood-Greenville
Harrisburg - LNCSTR-LEB-York	Houston	Huntsville-Decatur, Florida
Jonesboro	Joplin-Pittsburg	Knoxville
Lake Charles	Lexington	Little Rock - Pine Bluff
Lubbock	Memphis	Monroe - El Dorado
Montgomery	New York	Odessa - Midland
Oklahoma City	Panama City	Parkersburg
Philadelphia	Raleigh-Durham	Roanoke-Lynchburg
San Angelo	San Antonio	Sherman - ADA
Springfield, MO	Tampa-St.Pete, Sarasota	Tulsa
Wheeling-Steubenville	Wilkes Barre-Scranton	

TABLE A.2. Additional information on demographic covariates, shown in Figures B.1 and B.2

Covariate	Details
Age	Age of the head of household
White	Indicator variable if head of household is white
Asian	Indicator variable if head of household is Asian
Black	Indicator variable if head of household is African-American
College	Indicator variable if head of household is a college graduate
Democrat	Indicator variable if head of household is a Democrat (from party of registration)
Republican	Indicator variable if head of household is a Republican (from party of registration)
Has Children	Indicator variable if household has children
HH size	Number of individuals in the household
Hispanic	Indicator variable if head of household is Hispanic
Home Size (SF)	Size of home in square feet
Home Value	Home value of the household
Income	Household income
Male	Indicator variable if head of household is male
Married	Indicator variable whether household contains a married couple
No. Adults	Number of adults in the household
No. cars	Number of owned cars in the household
No. children	Number of children in the household
No. generations	Number of different family generations living in the household
Net worth	Estimated net worth at the household level
Own Home	Indicator variable whether the home being lived in is owned
Vehicle Year	Year of newest owned vehicle

TABLE A.3. Comparison of mean demographics in FWM data with a nationally-representative sample (GfK/MediaMark Survey of the American Consumer, 2012 Doublebase Waves) of cable subscribers.

	Source	
	FWM	GfK
HH Composition		
HH Size	2.84	3.01
No. Adults	2.36	2.28
Married	0.69	0.53
Has Children	0.49	0.39
No. Children	0.69	0.74
Own Home	0.97	0.69
Housing		
Home Value	141.32	265.63
No. Cars	1.84	1.9
Auto		
Vehicle Year	2,005.99	2,005.38
Net Worth	88.86	343.95
Income / Wealth		
Income	58.42	78.72
Age	54.45	46.99
Education		
White	0.86	0.76
Race		
Asian	0.01	0.03
Black	0.07	0.13
Hispanic	0.07	0.13
College	0.38	0.49
Party ID		
Republican	0.2	0.21
Democrat	0.47	0.24

Notes: Figures for FWM are simple averages of all households in tracking data. For GfK, respondents are weighted by GfK's inverse sampling weights. Income, wealth, and housing values use different bins and top-coded values in GfK and FWM data and hence some difference may be attributable to binning choices. Party ID in FWM data is a combination of survey response and party registration in partisan registration states, with registration preferred where it is available. In GfK it is based on survey response only.

Appendix B. Further Information on Balance Tests, Parallel Trends and Inference

We begin with two representations of the results on the balance of covariates.

FIGURE B.1. Balance Test - Treatment Effects on each Covariate

Treatment 'Effects' on Demographic Variables
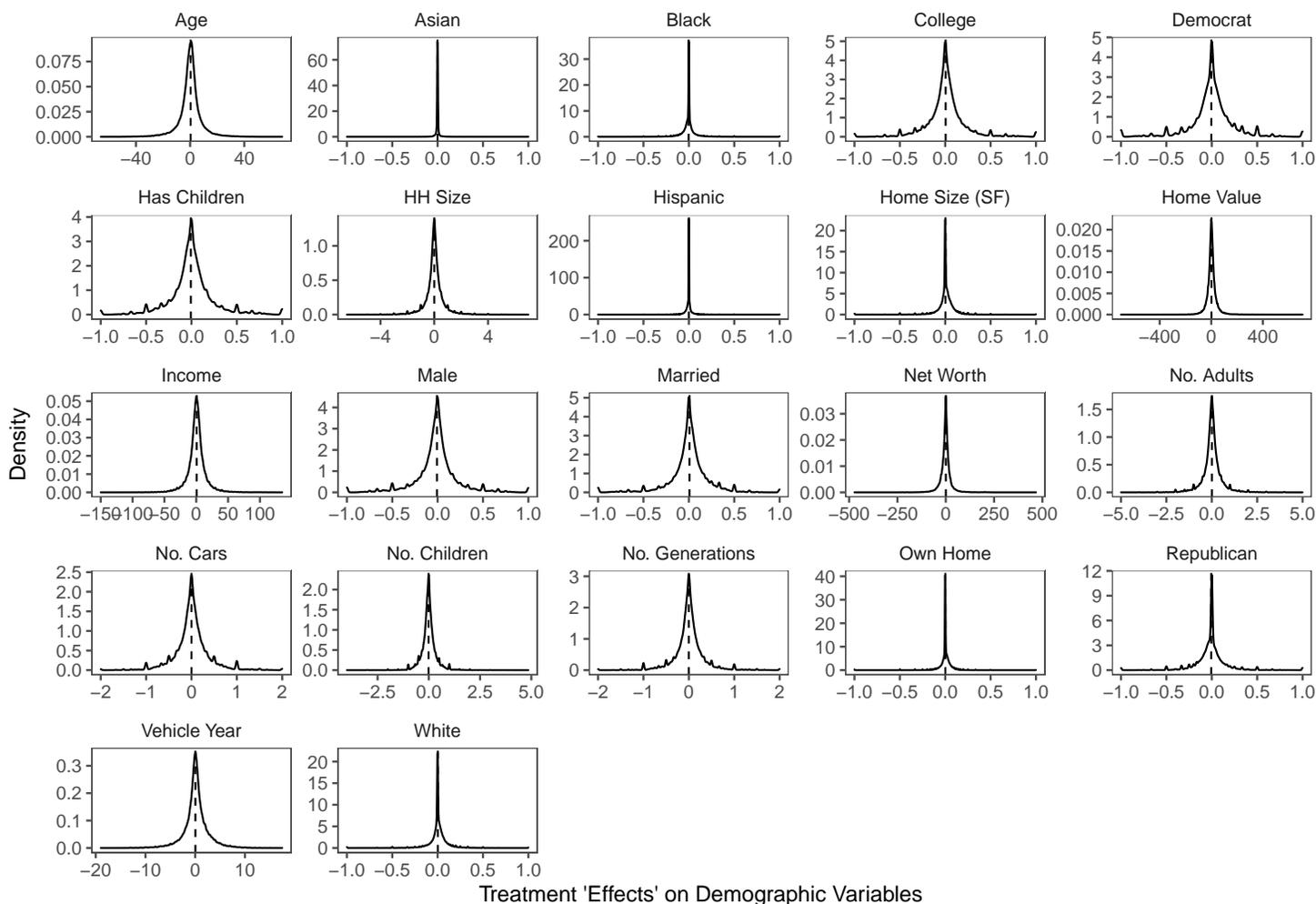
Notes: We present the distribution of estimated treatment effects for different covariates. Estimates are constructed by regressing each indicated covariate on an indicator for treatment status $T_{i,a}$. There is one such regression for every covariate and every ad in the sample. Our data is well balanced as the covariates are not significantly different across those groups.

FIGURE B.2. Balance Test - p-values of Treatment Effects on each Covariate
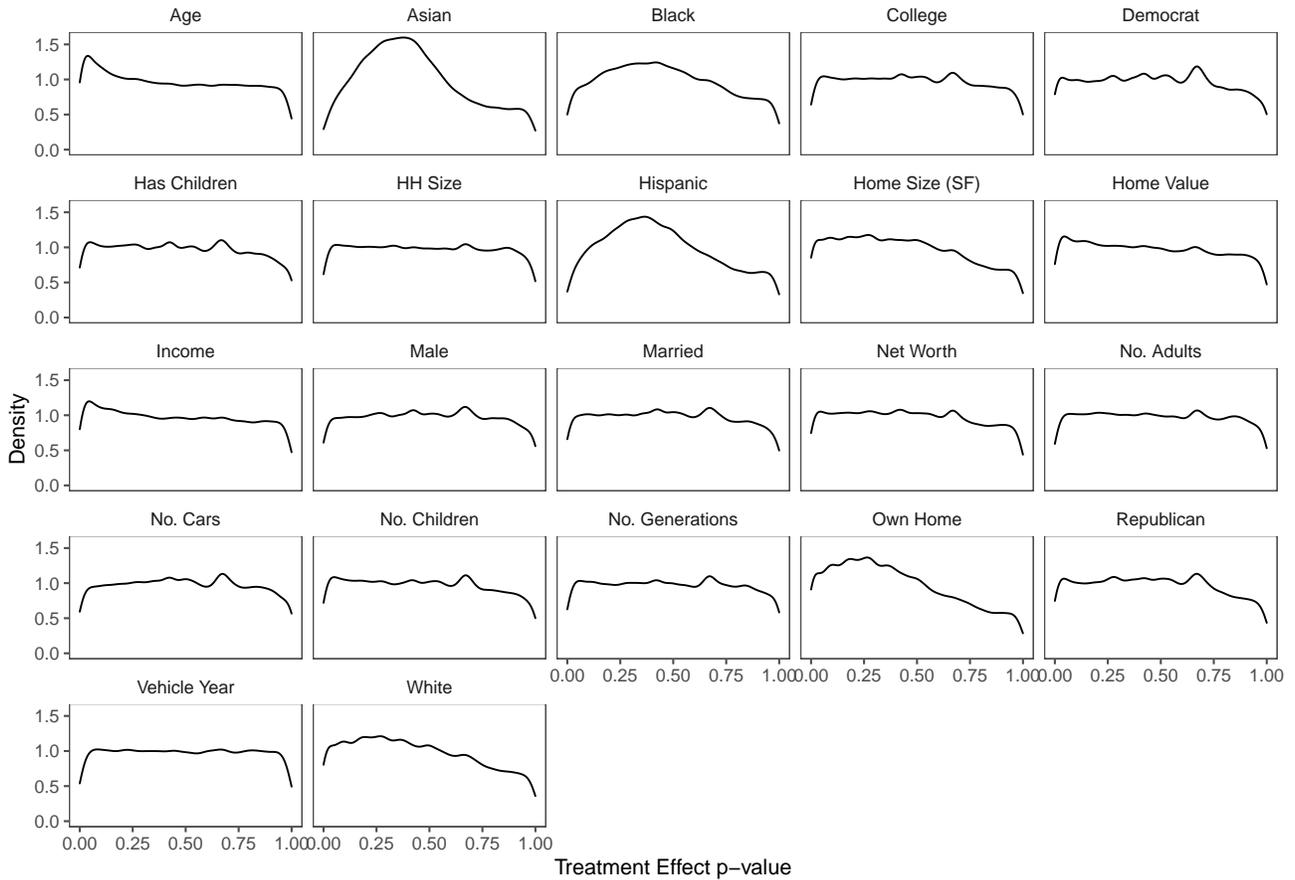
Notes: We present the distribution of p-values for different covariates under the null that the treatment effect of that covariate is equal to 0. The p-values should follow a uniform distribution if the null hypothesis that the treatment effect is 0 is true. We can see that, in general, our data is well balanced as the covariates are not significantly different across treatment and control. This graph complements Figure B.1.

Figure B.3 below shows that treatment and control groups are balanced in their exposure to *past* ads. Meanwhile, Figure B.4 shows that the distribution of past ad exposure to devices in the sample is stable over the last month of the campaign.

FIGURE B.3. Balance of Device-Level Assignments to Treatment/Control
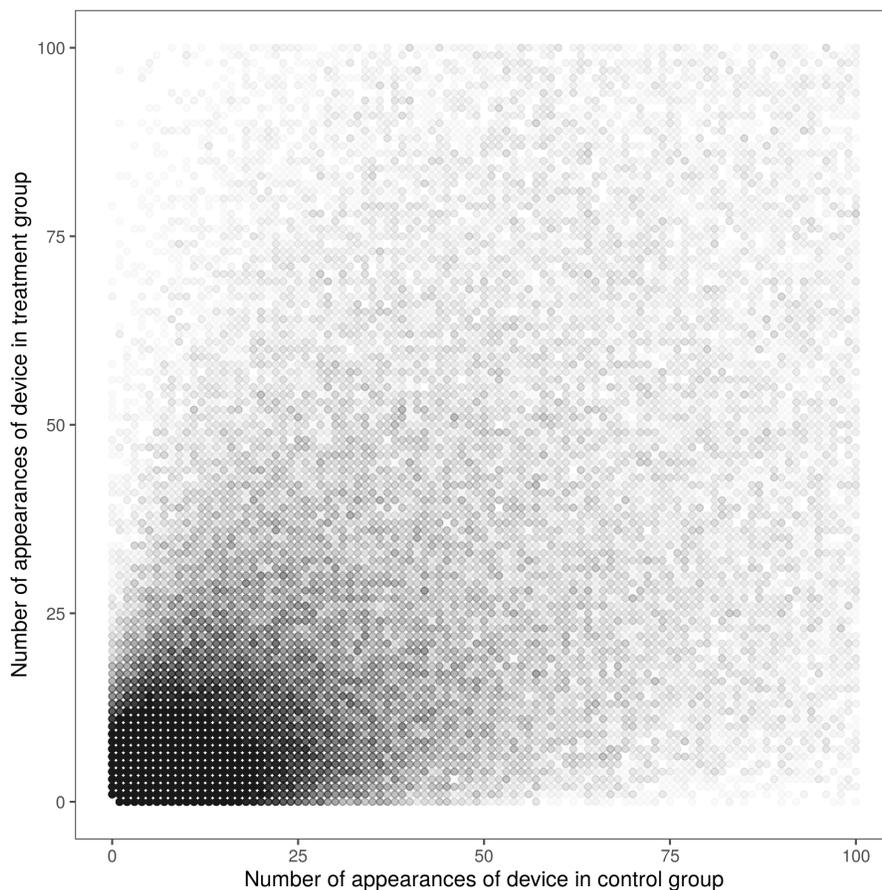
Notes: We present a scatter plot of the device-level counts of treatment/control group membership over the whole sample period. Each point is a device, with the number of times that a device appears in the control group is on the x-axis and the number of times that device appears in treatment is on the y-axis. This shows that treated viewers and control viewers are well balanced on the amount of past treatments.

FIGURE B.4. Balance of Device-Level Assignments to Treatment over Time
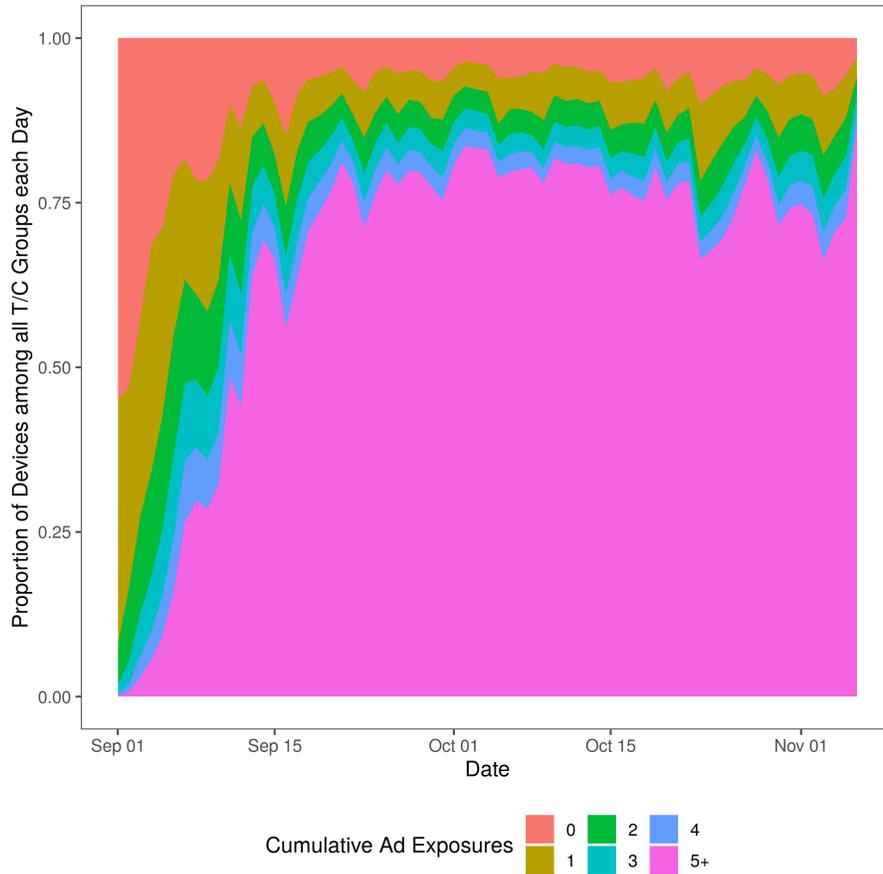
Notes: We show the distribution of cumulative exposure to political ads for devices in our estimation sample, and its evolution over time. For example, the height of the red area is the proportion of all devices in all treatment and control groups on a given day that have 0 prior exposures to political ads. While Figure B.3 shows that the amount of past exposure to ads over time is balanced across treated and control groups, this figure shows that the proportions of treated and control groups with different levels of cumulative exposure is stable over the last month and a half of the campaign.

B.1. Windowing choice and robustness to parallel trends violations

We use 24 hour windows before and after the ad to account for time patterns in news viewing. Our results use the difference in total viewing time in the window after ad time relative to that in the symmetric window before ad time. In the main text, we described a sufficient parallel pre-trends assumption for this context. However, we can rely on a weaker assumption for identification: we can allow differential trends across groups as long as any differential trend is cyclical within the 24 hour window. Such cyclical differential trends might arise if, for example, the "treated" viewers for an ad airing at 5:25PM are those who regularly watch news at 6 and are thus just beginning to watch news, whereas the "control" viewers are those who regularly watch at 5:30 and thus are about to stop watching. The 24-hour-window difference-in-differences estimate eliminates bias of this form, so long as this time pattern is consistent day to day. We now clarify and expand on this point.

Let $h \in \{\ldots, -2, -1, 1, 2, \ldots\}$ define hour windows relative to ad a. Specify a model of viewing at the hourly level as:

$$(B.1) \qquad y_{i,h} = \xi_h + T_i \psi_h + \delta_{a,h} T_i \mathbf{1}(h > 0) + \epsilon_{i,h},$$

where $y_{i,h}$ is news viewing by device i in hour window h, ξ_h represents hourly averages (relative to ad time), ψ_h the differential trend (relative to ad time) in hour h for the treatment group and $\delta_{a,h}$ the true treatment effect on each future hour.

A sufficient condition for identification of the (sum of) treatment effects in this specification is that $\psi_{+1} = \psi_{-24}, \psi_{+2} = \psi_{-23}$, and so on; i.e. that differential trends cycle on a 24 hour schedule. This can be seen by taking expectations of our 24 hour difference-in-differences estimator:

$$E[DD_H] = \sum_{h=1}^{H} \delta_{a,h} + \left(\sum_{h=1}^{H} \psi_h - \sum_{h'=-1}^{-H} \psi_{h'} \right),$$

where H is the window length in hours. Figure 3a in the main text shows that this assumption appears to hold in our data if we choose a 24 hour window length ($H = 24$).

We can also make use of Equation B.1 to decompose the 24 hour effect into hour-by-hour components, e.g. to recover the hourly treatment effects $\delta_{a,h}$. Given (B.1), note that the difference of differences

$$\frac{1}{N_{a,T}} \sum_{i:\ T_{i,a}=1} \left(y_{i,h} - y_{i,h-25}\right) - \frac{1}{N_{a,C}} \sum_{i:\ T_{i,a}=0} \left(y_{i,h} - y_{i,h-25}\right)$$

converges in probability to $\delta_{a,h} + \psi_h - \psi_{25-h}$ under standard regularity conditions and is thus a consistent estimate of the hourly effect under our 24 hour cyclicality assumption. Figure 4 in the main text shows the results of this decomposition, showing that the effect is concentrated in the hour following ad time, and has decayed to zero after about 6 hours (with the exception of a small rebound in hour 23, which is the hour prior to the ad time on the next day).

B.2. The bootstrap approach for computing standard errors in this design

While we estimate individual effects δ_a, we do not perform inference on each individual parameter: there are over 100,000 of them. We are instead interested in distributions of those effects. These include equally weighted and weighted averages of δ_a across all ads, or averages over subsets of ads (e.g. ads sponsored by a given party, ads sponsored by candidates in local or national races, ads within markets with a higher share of higher educated viewers) — see Table 2.

We estimated those effects using a plug-in approach (i.e. we plugged-in $\hat{\delta}_a$ for δ_a), and averaged over the number of ads *which we assume is fixed*. To perform inference, we then repeatedly resample with replacement from the stacked vector of $\hat{\delta}_a$, for the subset of ads we wished to study.

It is straightforward to find conditions for the validity of the bootstrap for those statistics. Recall that under standard regularity conditions, each difference-in-differences estimator $\hat{\delta}_a$ is consistent for δ_a and asymptotically normal (see Athey and Imbens, 2006). The validity of the bootstrap then follows from standard results on smooth and well behaved estimators. Since our parameters of interest are linear transformations of subsets of $\{\delta_a\}_a$, the plug-in

estimators for our weighted averages preserve the consistency and asymptotic normality results *under fixed a asymptotics* and ads a being independent. This is because consistency and asymptotic normality is preserved in the sum of independent random variables that are normally distributed.[43] The validity of the bootstrap follows from its consistency for each ad a together with the Continuous Mapping Theorem to aggregating them.

This argument can be generalized when the ad-level effects are correlated and, in fact, is done so in Athey and Imbens (2006), Section 6. This simply requires that the statistics $\hat{\delta}_a - \delta_a$ are jointly normal across a, which holds under standard regularity conditions and a Multivariate Central Limit Theorem.

---

[43]If a is allowed to increase with sample size n we obtain a nonstandard asymptotic: the weights applied to aggregate our statistics would also be changing. In this case, we would have to control the estimation error of those weights so that they disappear at a fast enough rate. This is beyond the scope of the current paper.

Appendix C. Placebo Tests and Robustness Checks

C.1. Deconstructing the Control Group

To investigate the possible influence of differential trends on our baseline results, we run our main specification on three distinct subsets of the control group which might be expected to have different news viewing trends. We split the control group for each ad into three groups: viewers who (i) tuned into the ad channel within 20 minutes of the ad airing, but tuned out before the ad came on, (ii) tuned in within 20 minutes after the ad aired, and (iii) were watching the channel the ad aired within 20 minutes of the ad airing, tuned out before the ad came on, and then tuned back in later. The difference in these groups' timing of viewing the program on which the ad ran might correlate with differences in news viewing trends if, for example, the tune-out-before group is ramping down its news viewing and the tune-in-after group is ramping up.

We present results using each subset of the control group defined above in Table C.1. It is clear that the qualitative result is robust to using any of these sub-groups alone: the estimate in all cases is positive and both substantively and statistically significant. There is some variation in the size of the estimated effect, however, with a range of about 2 minutes between the group with the largest estimated effect (the tuned-out-before group, C1) and that with the smallest (the tuned-in-after group, C2). The group which both tuned out before and tuned in after (C3) is intermediate between these two. If parallel trends held uniformly across the control group, there should be no expected difference in the estimates as all measure the same treatment effect. Hence, the range of variation in the estimates (2 minutes) is informative about the degree to which violation of parallel trends could bias the estimate upwards.

TABLE C.1. Average Effects and Bootstrapped Confidence Intervals, by Control Group Split.

	Inv. Variance	Effect (Minutes) Total Devices	Equal Weight
	(1)	(2)	(3)
C1	4.617	4.366	2.203
	(4.369, 4.860)	(4.126, 4.601)	(1.862, 2.544)
C2	2.652	2.576	2.663
	(2.440, 2.861)	(2.370, 2.781)	(2.274, 3.019)
C3	4.277	4.109	2.890
	(3.965, 4.553)	(3.801, 4.387)	(2.439, 3.384)

Notes: The table reports average effects, weighted by 1) the geometric mean of treatment and control group size, 2) total devices in treatment and control groups and 3) a simple equally weighted average. Confidence intervals are the central 95% interval of 1000 bootstrap replicates within each subgroup. Group C1 is devices in control group which tuned out prior to ad time; C2 is devices in control group which tuned in after ad time; and C3 is devices in control group who tuned out before ad time, and then tuned back in after ad time.

C.2. Placebo Tests

This section presents the results for the placebo tests analyzed in Section 4.3. The first two tests re-estimate our main specifications using viewership of two classes of non-news programs (sports and comedy) instead of news viewing following the airing of the ad. We also run the split-control specifications described in Appendix C.1 on these outcomes. The results are presented in Tables C.2-C.3.

Altogether, the results support our main specification. We refer the reader to our analysis in Section 4.3. The next three sections present a series of robustness checks.

TABLE C.2. Average Effects and Bootstrapped Confidence Intervals, Comedy viewing placebo.

| | Effect (Comedy Viewing Minutes) | | |
| | Inv. Variance | Total Devices | Equally Weighted |
	(1)	(2)	(3)
All	−0.999	−0.973	−1.655
	(−1.110, −0.891)	(−1.087, −0.865)	(−1.911, −1.418)
C1	−2.547	−2.507	−2.911
	(−2.715, −2.375)	(−2.678, −2.336)	(−3.220, −2.644)
C2	0.588	0.591	−0.127
	(0.406, 0.760)	(0.412, 0.767)	(−0.432, 0.178)
C3	−0.310	−0.296	−1.018
	(−0.543, −0.079)	(−0.529, −0.073)	(−1.392, −0.633)

Notes: The table reports average effects on comedy program viewing (in minutes), weighted by 1) the geometric mean of treatment and control group size, 2) total devices in treatment and control groups and 3) a simple equally weighted average. Confidence intervals are the central 95% interval of 1000 bootstrap replicates within each subgroup. Group C1 is devices in control group which tuned out prior to ad time; C2 is devices in control group which tuned in after ad time; and C3 is devices in control group who tuned out before ad time, and then tuned back in after ad time. 'All' row includes all three types as controls.

TABLE C.3. Average Effects and Bootstrapped Confidence Intervals, Sports viewing placebo.

| | Effect (Sports Viewing Minutes) | | |
| | Inv. Variance | Total Devices | Equally Weighted |
	(1)	(2)	(3)
All	−0.434	−0.416	−1.409
	(−0.574, −0.287)	(−0.563, −0.264)	(−1.668, −1.152)
C1	−1.429	−1.442	−2.456
	(−1.644, −1.210)	(−1.657, −1.225)	(−2.752, −2.157)
C2	0.378	0.395	−0.289
	(0.189, 0.565)	(0.201, 0.585)	(−0.608, 0.039)
C3	0.237	0.323	−0.306
	(−0.030, 0.526)	(0.051, 0.612)	(−0.716, 0.091)

Notes: The table reports average effects on sports program viewing (in minutes), weighted by 1) the geometric mean of treatment and control group size, 2) total devices in treatment and control groups and 3) a simple equally weighted average. Confidence intervals are the central 95% interval of 1000 bootstrap replicates within each subgroup. Group C1 is devices in control group which tuned out prior to ad time; C2 is devices in control group which tuned in after ad time; and C3 is devices in control group who tuned out before ad time, and then tuned back in after ad time. 'All' row includes all three types as controls.

C.3. Robustness Checks

Our baseline difference-in-differences results assumed a 20 minute window defining treatment (i.e. households were considered "subjects" if they had tuned-in to the channel where the ad was aired within 20 minutes of the ad). Below, we check the robustness of our estimates if this window is reduced to 5 and 10 minutes.

Tables C.4-C.5 show that our main specification results in Table 2 still hold with these different definitions. In fact, we show that with smaller windows, the differences in estimates across the deconstructed control groups decreases relative to the baseline results. In Tables C.6-C.7 we further confirm the sponsor heterogeneity findings from Table 2 discussed in the main text.

TABLE C.4. Average Effects and Bootstrapped Confidence Intervals, by Control Group Split - 5 Minute Window

| | Effect (Minutes) | | |
| | Inv. Variance | Total Devices | Equal Weight |
	(1)	(2)	(3)
C1	4.248	3.972	1.830
	(3.936, 4.553)	(3.678, 4.267)	(1.370, 2.292)
C2	3.592	3.436	2.836
	(3.338, 3.849)	(3.189, 3.683)	(2.354, 3.328)
C3	5.220	5.010	3.726
	(4.744, 5.706)	(4.548, 5.492)	(3.093, 4.392)

Notes: The table reports average effects, weighted by 1) the geometric mean of treatment and control group size, 2) total devices in treatment and control groups and 3) a simple equally weighted average. Confidence intervals are the central 95% interval of 1000 bootstrap replicates within each subgroup. Group C1 is devices in control group which tuned out up to 5 minutes prior to ad time; C2 is devices in control group which tuned in up to 5 minutes after ad time; and C3 is devices in control group who tuned out before ad time, and then tuned back in after ad time.

TABLE C.5. Average Effects and Bootstrapped Confidence Intervals, by Control Group Split - 10 Minute Window.

		Effect (Minutes)	
	Inv. Variance	Total Devices	Equal Weight
	(1)	(2)	(3)
C1	4.800	4.501	2.096
	(4.529, 5.065)	(4.239, 4.759)	(1.677, 2.535)
C2	3.617	3.480	2.773
	(3.407, 3.851)	(3.271, 3.709)	(2.358, 3.181)
C3	5.116	4.877	3.324
	(4.762, 5.462)	(4.537, 5.210)	(2.808, 3.847)

Notes: The table reports average effects, weighted by 1) the geometric mean of treatment and control group size, 2) total devices in treatment and control groups and 3) a simple equally weighted average. Confidence intervals are the central 95% interval of 1000 bootstrap replicates within each subgroup. Group C1 is devices in control group which tuned out up to 10 minutes prior to ad time; C2 is devices in control group which tuned in up to 10 minutes after ad time; and C3 is devices in control group who tuned out before ad time, and then tuned back in after ad time.

TABLE C.6. Average Effects and Bootstrapped Confidence Intervals, by Subgroup - 5 Minute Window.

	Subgroup		Effect (Minutes)	
		Inv. Variance	Total Devices	Equal Weight
		(1)	(2)	(3)
Full Sample		4.051	3.845	2.754
		(3.840, 4.272)	(3.631, 4.060)	(2.310, 3.191)
Sponsor	House	3.571	3.469	1.496
		(3.137, 3.976)	(3.040, 3.875)	(0.704, 2.282)
	President	3.559	3.308	2.819
		(3.171, 3.942)	(2.935, 3.684)	(2.022, 3.632)
	Senate	4.879	4.588	3.837
		(4.264, 5.502)	(3.979, 5.207)	(2.699, 4.903)
	Statewide	6.701	6.612	4.217
		(6.104, 7.296)	(6.024, 7.221)	(3.131, 5.303)
	Outside Group	2.735	2.523	2.472
		(2.212, 3.224)	(2.002, 3.009)	(1.551, 3.452)
Incumbency	Challenger	5.164	5.002	4.108
		(4.806, 5.541)	(4.640, 5.371)	(3.362, 4.848)
	Incumbent	3.773	3.558	1.933
		(3.466, 4.072)	(3.257, 3.853)	(1.319, 2.509)

Notes: The table reports average effects, weighted by 1) the geometric mean of treatment and control group size, 2) total devices in treatment and control groups and 3) a simple equally weighted average. Confidence intervals are the central 95% interval of 1000 bootstrap replicates within each subgroup.

TABLE C.7. Average Effects and Bootstrapped Confidence Intervals, by Subgroup - 10 Minute Window.

| | Subgroup | Effect (Minutes) | | |
| | | Inv. Variance | Total Devices | Equal Weight |
		(1)	(2)	(3)
Full Sample		4.431	4.207	2.729
		(4.248, 4.607)	(4.030, 4.384)	(2.376, 3.086)
Sponsor	House	3.613	3.492	0.681
		(3.250, 3.968)	(3.142, 3.841)	(−0.043, 1.414)
	President	4.222	3.927	3.506
		(3.885, 4.570)	(3.597, 4.273)	(2.786, 4.211)
	Senate	5.247	4.950	3.829
		(4.717, 5.739)	(4.426, 5.437)	(2.883, 4.831)
	Statewide	7.262	7.187	4.243
		(6.767, 7.748)	(6.689, 7.694)	(3.387, 5.149)
	Outside Group	2.920	2.698	2.223
		(2.463, 3.343)	(2.258, 3.126)	(1.423, 3.060)
Incumbency	Challenger	5.577	5.393	3.690
		(5.248, 5.883)	(5.065, 5.700)	(3.050, 4.325)
	Incumbent	4.189	3.959	2.291
		(3.949, 4.451)	(3.713, 4.218)	(1.789, 2.796)

Notes: The table reports average effects, weighted by 1) the geometric mean of treatment and control group size, 2) total devices in treatment and control groups and 3) a simple equally weighted average. Confidence intervals are the central 95% interval of 1000 bootstrap replicates within each subgroup.

As shown in Figure B.2, one concern with our empirical specification could be that some variables, for example Age and Hispanic, appear unbalanced across a larger share of treatment and control groups than what should be expected. To check the robustness of our results from Table 2 to this dimension, we run our main specifications again on the subset of ads (treatments) for which the p-value for treatment balance is larger than 0.05. The results are shown below in Tables C.8-C.9. We can see that our main conclusions remain unchanged.

TABLE C.8. Average Effects and Bootstrapped Confidence Intervals: Age-balanced ads only.

	Inv. Variance	Effect (Minutes) Total Devices	Equally Weighted
	(1)	(2)	(3)
Treated	6.176	6.063	4.374
	(5.944, 6.407)	(5.831, 6.294)	(4.143, 4.606)

Notes: The table reports average effects, weighted by 1) the geometric mean of treatment and control group size, 2) total devices in treatment and control groups and 3) a simple unweighted average. Confidence intervals are the central 95% interval of 1000 bootstrap replicates. The sample is restricted to the set of ads with p-value greater than 0.05 for the hypothesis that treatment and control groups have no difference in mean age.

TABLE C.9. Average Effects and Bootstrapped Confidence Intervals: Ads balanced on Hispanic share only.

	Inv. Variance	Effect (Minutes) Total Devices	Equally Weighted
	(1)	(2)	(3)
Treated	5.947	5.847	3.861
	(5.749, 6.146)	(5.648, 6.045)	(3.662, 4.060)

Notes: The table reports average effects, weighted by 1) the geometric mean of treatment and control group size, 2) total devices in treatment and control groups and 3) a simple unweighted average. Confidence intervals are the central 95% interval of 1000 bootstrap replicates. The sample is restricted to the set of ads with p-value greater than 0.05 for the hypothesis that treatment and control groups have no difference in the fraction of Hispanic viewers.

C.4. Regression Discontinuity Design (RDD) Estimates

Below we provide the formal estimates for the specifications described in Section 4.4.

TABLE C.10. Regression Discontinuity Design Estimates.

	Effect (Seconds)			
	(1)	(2)	(3)	(4)
Local Linear Estimates				
Treated	89.77	111.84	228.20	139.22
	(44.80)	(23.86)	(12.26)	(19.02)
	[0.011]	[0.059]	[0.000]	[0.000]
Local Quadratic Estimates				
Treated	168.56	63.94	147.48	138.30
	(65.95)	(33.93)	(16.95)	(18.50)
	[0.416]	[0.116]	[0.000]	[0.000]
Bandwidth	15 seconds	1 minute	5 minutes	CCT optimal

Notes: The table reports estimated effects from a regression discontinuity design with robust-bias corrected estimates following Calonico et al. (2014). Each column uses a certain bandwidth, ranging from 15 seconds to 5 minutes. The CCT optimal bandwidth is given by the one from Calonico et al. (2014), and is equal to 101.69 seconds for the local linear specification, and 242.94 seconds for the local quadratic specification. Standard errors in parentheses, p-values from robust and bias-corrected inference in brackets.

Appendix D. Missing Programming Data

As mentioned in the main text, one limitation of our data is that the program database (Fourthwall) is incomplete, an issue that is most pronounced for local network affiliates in certain DMAs, primarily in the West. National cable channels have nearly complete coverage. The national cable channels have an average of roughly 30 news programs per day in our data, and have some program entry in all 154 days of our coverage period. The local affiliates average about 9.6 news programs per day on days with coverage, but almost 70% of total

station-days are missing. There are 71 stations in 15 DMAs (out of 663 and 60 respectively) that are largely unaffected by schedule missingness. These coincide with markets with the largest number of subscribers in our data. We restrict our analysis to this subset of DMAs in the robustness analysis below.

While solutions to this issue are limited by the data availability, we do not believe it poses a significant threat to our analysis, for several reasons.

First, this missingness cannot produce bias away from zero in the estimates in our set-up. This is because the variation we use is within media markets (i.e. within the same program schedules). Hence, the measurement error induced by such incompleteness cannot be correlated with treatment group status. It is possible that missing program data generates bias towards zero, as we are effectively plugging in zeros for the outcome variable on days with missing schedule data. In fact, our estimates rise when we include only markets with relatively complete schedule data; see Table D.1 below.

Regarding the efficiency of our estimates, which would be affected by this incompleteness: our results in Table 2 are reasonably similar whether weighted by sample size or equally weighted. This is suggestive that average estimates are similar across ads with smaller and larger number of viewers within media markets. Hence, data incompleteness does not appear to affect the efficiency of our estimates in an economically meaningful way when this missingness is correlated with viewership numbers within markets.

However, data incompleteness could be correlated to the magnitude of effects in media markets. For example, data incompleteness could be higher in markets where ads are less effective. This does not appear to be the case: when we drop media markets with the most missing data in our sample and re-estimate the specifications in Table 2, we find even larger results than before. These are shown in Table D.1 below.

TABLE D.1. Average Effects and Bootstrapped Confidence Intervals: Dropping DMAs with missing local news program data.

| | Effect (Minutes) | | |
| | Inv. Variance | Total Devices | Equally Weighted |
	(1)	(2)	(3)
Treated	8.015	7.913	10.325
	(7.826, 8.205)	(7.724, 8.103)	(10.135, 10.515)

Table reports average effects, weighted by 1) the geometric mean of treatment and control group size, 2) total devices in treatment and control groups and 3) a simple equally weighted average. Confidence intervals are the central 95% interval of 1000 bootstrap replicates. Sample drops ads run in DMAs with extensive missingness of the local news program schedule.

Appendix E. Additional Results on Tune-out

Interpreting tune-out frequencies requires an appropriately constructed baseline for comparison. Just as a post-treatment increase in an outcome cannot be interpreted as a causal effect without reference to the counterfactual change estimated by the corresponding increase in an untreated control group, the rate of tune-out for political ads only has meaning in relation to the comparable rate for other content.

We construct an ad-level baseline rate for comparison by the following procedure. For every device tuned in at the instant an ad aired (i.e., the treatment group for some ad), we compute an indicator $o_{i,a}^A \in \{0, 1\}$ which is 1 if the device registered a tuning event prior to the end of the ad and 0 otherwise. We then use the device's viewing history to randomly select a device-specific time somewhere else in the same day that satisfies our criteria for inclusion in an ad treatment group (device is on, and has had tuning event in the previous 20 minutes).[44] We denote this randomly selected time by $r_{i,a}$ and measure an indicator $o_{i,a}^R \in \{0, 1\}$ which is 1 if the device registered a tuning event between $r_{i,a}$ and $r_{i,a} + 30$ and 0 otherwise. For each ad we construct average tune-out rates $O_a^A = \sum_{i\,:\,T_{i,a}=1} o_{i,a}^A/|T_{i,a} = 1|$ and $O_a^R = \sum_{i\,:\,T_{i,a}=1} o_{i,a}^R/|T_{i,a} = 1|$.

---

[44]We exclude the 30 minutes prior to the ad air time and the duration of the ad itself, and sample uniformly over all other seconds that satisfy the two stated criteria.

The difference between the two rates is a measure of the ad's (dis-)utility relative to the average utility of TV viewing[45], but most importantly, serves as a measure of compliance: how many people exposed to treatment persisted until the end. As such, it allows us to assess whether our estimates in the previous section could be different depending on the amount of exposure to advertisement.

Figure E.1 shows the distributions of tune-out rates for both actual ads and our randomly-selected comparison times. Each observation in the underlying dataset is a single ad; we plot the density of tune-out rates — the fraction of viewers who were tuned in at the beginning of the ad who tuned out before the end — weighted by the number of devices that were tuned in at the ad start time. The distribution for actual ads is compared to the distribution of the fraction of devices that tuned out within 30s of the device-specific comparable time. Comparing the two distributions shows that on average the tune out fractions are low, and very similar between ad times and randomly selected times: the mean tune-out fraction is 4.2% for ads and 4.6% for the randomly selected comparable times. Interestingly, though the means are very similar, the distribution for political ads is more polarized - there are more observations with very low tune-out rates but also more with very high tune-out rates.

---

[45]This measure can be interpreted as an (up-to scale) approximation of consumer surplus under small shares of tuning-out in a standard discrete choice model with Logit shocks. Additional details are available from the authors.

FIGURE E.1. Tune-Out Probability of Political Ads
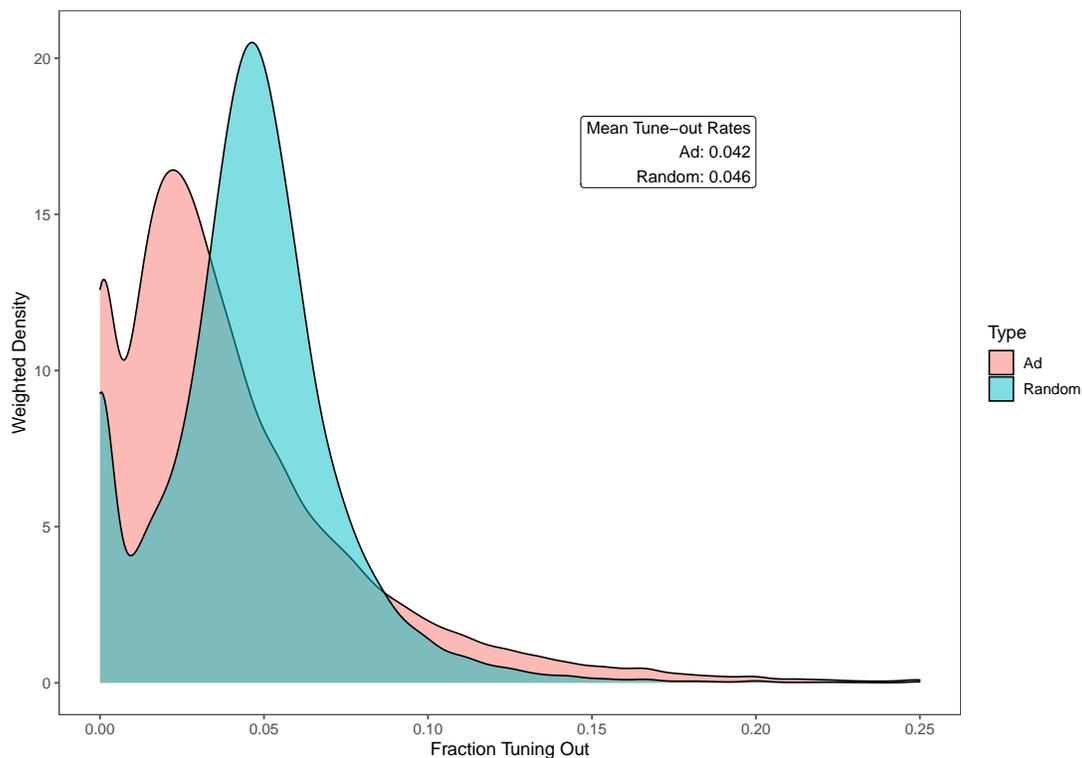
Notes: We present the distribution of estimates of the tune-out rate, the fraction of people who tuned out before the ad ends. We compare to the fraction who tuned out at a device-specific randomly selected time somewhere else in the day and that satisfies our criteria for treatment (device is on, and has had tuning event in the previous 20 min). This generates a rate of tune out for other content for comparison to the rate among political ads. The plot displays the density, weighted by treatment group size, of ad-level estimates of ad tune-out rate and randomly-selected-time tune-out rate.