

Social Media, News Consumption and Polarization: Evidence from a Field Experiment

Ro'ee Levy*

October 1, 2019

[Click here for the latest version](#)

Abstract

Does the rise of social media increase the consumption of news matching the consumer's ideology and lead to greater polarization? I estimate the effects of social media news exposure by conducting a large field experiment that randomly offers participants subscriptions to conservative or liberal news outlets on Facebook. I collect data on the chain of media effects: subscriptions to outlets, exposure to news on Facebook, visits to online news sites, sharing of posts, and changes in political beliefs. Four main findings emerge. First, exogenous variation in the news individuals are exposed to on social media substantially affects the news sites they visit, suggesting that search costs play an important role in determining the news that individuals consume. Second, exposure to counter-attitudinal news decreases affective polarization, defined as negative attitudes toward the opposing political party. Third, in contrast to the effect on attitudes, I find no evidence that the political leaning of news consumed affects political opinions. Fourth, I decompose the channels increasing exposure to pro-attitudinal news on social media and find that Facebook's algorithm is less likely to supply individuals with posts from counter-attitudinal outlets (a "filter bubble"). Together, these results imply that social media algorithms may be limiting exposure to counter-attitudinal news and thus increasing polarization.

*Yale University, roee.levy@yale.edu. I am deeply grateful to my advisors Dean Karlan, Joseph Shapiro and Ebonya Washington for their guidance throughout this project. I thank Dirk Bergemann, Leonardo Bursztyn, Alex Coppock, Oren Danieli, Eduardo Fraga, Martin Mattsson, David Rand, Oren Sarig, David Schonholzer, Katherine Wagner and Mor Zoran for helpful comments. I also thank seminar participants at Berkeley, Northwestern and Yale for their helpful comments. Financial support from the ISPS Field Experiment's Initiative, the Tobin Center for Economic Policy, the Yale Program in Applied Economics and Policy, and David Rand is greatly appreciated. All errors are my own. The experiment is registered at the AEA RCT registry (ID 0002713).

In 2018, two-thirds of American adults consumed news on social media, compared to fewer than one in eight Americans in 2008. Based on Pew surveys, Facebook is the most dominant social media platform for news consumption (Shearer and Matsa, 2018), and “among millennials, Facebook is far and away the most common source for news about government and politics” (Pew, 2014). As social media becomes a major news source, there is growing apprehension over its effects on public opinion. A primary concern is that due to the unique features of social media, individuals are exposed to more pro-attitudinal news, defined as news matching their ideology, and as a result polarization increases (Sunstein, 2017). Furthermore, several recent scandals, including the Russian-based campaign to influence the 2016 US presidential elections, the spread of fake news on social media, and Cambridge Analytica’s attempt to affect political races, highlight a concern that individuals may be targeted and easily manipulated on social media.

In this paper, I ask two questions. First, how does social media affect news consumption habits? Second, what is the effect of social media news consumption on political opinions and affective polarization, defined as negative attitudes toward the opposing party? I study these questions by conducting a large online field experiment randomizing exposure to news outlets on social media and collecting survey, browsing, and social media data. I find that social media algorithms limit exposure to counter-attitudinal news (i.e., limit the number of posts from counter-attitudinal outlets observed in the social media feed), news exposure on social media substantially affects the news sites individuals visit, and exposure to counter-attitudinal news decreases affective polarization.

To motivate the experiment, I first test if news consumed through social media is different than other online news and focus specifically on Facebook. I merge data on browsing behavior, voting, and news outlets to show that news sites visited through Facebook tend to better match the consumer’s ideology, compared to other news sites visited. Using data from the experiment’s control group, I distinguish between two main factors contributing to the consumption of pro-attitudinal news through social media: links shared by one’s social network (i.e., Facebook friends), and link shared by the wide variety of free Facebook pages, which are organizations individuals can subscribe to when personalizing their social media feed (including virtually all media outlets). While the former factor, described as an “echo chamber” is often discussed in the popular press, I find that the latter factor, subscriptions to Facebook pages, is mostly driving the increased visits to pro-attitudinal websites through Facebook. The experiment creates variation in these subscriptions.

I recruited American adult Facebook users to the experiment using Facebook ads. Participants completed a baseline survey and at the end of the survey they were randomly assigned to a liberal treatment, a conservative treatment or a control group. Participants in the liberal and conservative treatments were offered subscriptions to four liberal or four conservative news outlets on Facebook (such as MSNBC or Fox News). When participants subscribe to an outlet by “liking” its Facebook page,¹ posts shared by the outlet could subsequently appear in their social media feed.

¹To subscribe to content from an organization, the user “likes” the organization’s page on Facebook. To simplify

Participants exposed to the posts could view headlines directly on Facebook and they could click on links in the posts to consume the full news stories in the outlets' websites.

I designed the experiment to have high external validity. Besides the initial offer to subscribe to outlets, which is common on social media, the intervention did not directly intervene in any behavior. The news supplied to participants was the actual news provided by leading media outlets during the study period. Facebook's algorithm determined which of the posts shared by the subscribed outlets appeared in the participants' Facebook feeds. Finally, participants decided whether to read, skip, or share specific posts. As a result, the treatment is almost identical to the experience of millions of Americans who subscribe to news outlets on Facebook.

Approximately two months after the baseline survey, participants were invited to an endline survey to measure the effect of the treatment on their political opinions and attitudes. Besides political opinions, my main outcome of interest is affective polarization since there is agreement that this measure of polarization has been increasing over time and concern over its implications on governance, accountability of elected officials and even labor markets.² When analyzing the effect on affective polarization, I re-define the treatments as a pro-attitudinal treatment and a counter-attitudinal treatment, based on the consumer's baseline ideology.³ The outcomes and specifications used to analyze the effect on political opinions and affective polarization were pre-registered.⁴

I ask participants to log in to the survey using their Facebook account and collect data on the participants' subscriptions to outlets on Facebook along with data on the posts shared by most participants.⁵ In addition, a browser extension I developed for this paper provides data on the news a subset of participants were exposed to on Facebook and the news sites they visited.

In both the liberal and conservative treatments, approximately half of the participants complied with the intervention by subscribing to at least one outlet. Interestingly, even among participants assigned to a counter-attitudinal treatment, the compliance rate was 45%. The intervention increased exposure to posts from the offered outlets and, as a result, substantially affected the mean slant (the media bias) of participants' Facebook feeds. The combined treatment-on-the-treated (TOT) effects of the liberal and conservative treatments equal approximately half of the gap between the slant of the feed of liberals and conservatives.

This paper has four main findings. First, exposure to news on social media substantially affects online news consumption. As a result of the treatment's effect on the participants' social media

terminology, throughout the paper I will describe the action of liking the page of a news organization as subscribing to an outlet on Facebook.

²Examples of papers showing that affective polarization has increased include Gentzkow (2016); Iyengar and Krupenkin (2018); Lelkes (2016). Iyengar et al. (2019) provide a recent review on affective polarization and its implications.

³A pro-attitudinal treatment is defined as a liberal treatment assigned to a liberal participant or a conservative treatment assigned to a conservative participant. A counter-attitudinal treatment is defined as a liberal treatment assigned to a conservative participant or a conservative treatment assigned to a liberal participant.

⁴The experiment was registered in the American Economic Association Randomized Registry for Randomized Controlled Trials under trial number 0002713.

⁵This data is only collected from participants who provide permissions to access this data.

feeds, participants visited the news sites of the outlets they were randomly offered, even when the outlets were counter-attitudinal. This implies that although social media is typically associated with pro-attitudinal news, individuals are willing to engage with counter-attitudinal news when it is accessible on social media. The effect of the social media feed on online news consumption is further driven by the fact that individuals do not re-optimize their browsing behavior to keep the slant of the news sites they visit constant, as both the liberal and conservative treatments had a significant effect on the mean slant of news sites visited. The difference between the TOT effects of the treatments is similar to the difference between the slant of news sites visited in New York and Alabama. This suggests that search costs play an important role when individuals decide which news to consume, as individuals shift their consumption habits when articles from specific outlets become more accessible such that the search costs for these articles decrease. Consequently, the algorithms determining which articles appear in social media feeds can drastically alter news consumption habits.

My second finding is that exposure to counter-attitudinal news *decreases* affective polarization, compared to pro-attitudinal news. I construct an affective polarization index measuring attitudes toward political parties. The measure includes questions such as how participants feel toward their own party and the other party (i.e. a “feeling thermometer”), and how they would feel if their son or daughter married a Democrat or Republican. I find that the counter-attitudinal treatment decreases the index by 0.03 standard deviations compared to the pro-attitudinal treatment.⁶ To compare the result to existing benchmarks, I focus on the feeling thermometer questions. The experiment’s ITT and TOT effects decreased the difference between participants’ feelings toward their party and the other party by 0.58 and 0.98 degrees on a 0-100 scale, respectively. For comparison, this measure of affective polarization increased by 3.83 degrees between 1996 and 2016.

To estimate the magnitude of the effect on polarization, I conduct back-of-the-envelope calculations measuring how counterfactual social media platforms would change affective polarization. I find that if individuals were exposed to an equal share of pro- and counter-attitudinal news on Facebook, the difference in the feeling thermometer questions would decrease by 3.76 degrees. This estimate should be interpreted cautiously since it assumes that the treatment affected attitudes only through the effect on the share of counter-attitudinal news, since it does not take into account general equilibrium effects and since it is based on the effect found over a two-month period. In a second counterfactual, based on similar assumptions, I calculate that if Facebook would become only slightly more balanced, such that news consumed through Facebook would be as congruent as news consumed through other means, the difference in the feeling thermometer would decrease by 1.11 degrees.

The paper’s third finding is that the slant of news that participants were exposed to on social media did not affect political opinions. The effect of the liberal and conservative treatments on a political opinion index, focusing on issues covered in the news during the study period, such as

⁶All estimates in the paper are intention-to-treat estimates (ITT) unless noted otherwise.

the March for Our Lives Movement or the Special Counsel investigation, is economically small, precisely estimated and is not statistically significant.

Why did the treatments affect attitudes toward political parties but not political opinions? I propose a simple model where news consumers have heterogeneous weights on beliefs that combine to form political opinions. An attitude of an individual toward a party is a function of the distance between the party's political opinion and the opinion it would form based on the individual's beliefs, weighted according to the party's weights. I show that the model is consistent with the results if the intervention affected beliefs on which the participants place low weights and the opposing party places high weights. Intuitively, the participants may have learned the logic behind some of the arguments made by a party and so they were able to rationalize the party's opinion, even if they continued to disagree with it.

The paper's fourth finding is that Facebook's algorithm limits exposure to counter-attitudinal content. I decompose the gap between exposure to posts from the pro- and counter-attitudinal outlets offered in the experiment into three main explanations: (1) participants are less likely to subscribe to counter-attitudinal outlets; (2) Facebook's algorithm is less likely to supply posts from counter-attitudinal outlets, conditional on participants subscribing to them; (3) participants use Facebook less frequently in the counter-attitudinal treatment. While I find evidence for all three forces, the most important explanation for the gap in exposure is Facebook's algorithm. The decomposition implies that social media algorithms supplying personalized sets of stories to consumers based on their perceived interests can substantially affect the type of news people are exposed to and increase exposure to news matching the consumer's ideology.

Combining the results paints a complicated picture. On the one hand, social media's algorithms limit exposure to counter attitudinal news. In turn, this affects the news sites individuals visit and increases affective polarization. The study also shows, however, that individuals are willing to engage with counter-attitudinal news and a subtle nudge can substantially diversify news consumption and decrease polarization.

This paper contributes to the literature on social media and online news consumption by showing *why* social media is associated with pro-attitudinal news consumption. In the economics literature, Gentzkow and Shapiro (2011) show that news consumption online is not more segregated than offline news, i.e., that the difference between the news consumption of liberals and conservatives is not greater when news is consumed online. Their study uses data from 2004-2009 when social media was still in its infancy. More recent papers both suggest that social media may be playing a role in increasing online segregation and argue that social media does not increase segregation and may even have a moderating influence (Barberá, 2015; Flaxman et al., 2016; Guess, 2018; Peterson et al., 2018). Most of the papers in this literature are based on browsing data, but they lack data on news individuals are exposed to on social media. As a result, these papers cannot test for segregation *within* one's social media feed. By matching unique data on the Facebook feed with browsing data, I show that social media is indeed associated with pro-attitudinal news

consumption and that the pages individuals choose to subscribe to on Facebook are driving the increase in pro-attitudinal news consumption.

I also use this data to provide the first experimental evidence that social media algorithms are increasing exposure to pro-attitudinal news. These algorithms have long been suspected to increase segregation in news consumption (Sunstein, 2017; Tufekci, 2015). However research on their effect is limited. Bakshy et al. (2015) analyze Facebook’s data to show that exposure to counter-attitudinal news is mostly limited by individual choices and not by algorithmic ranking.⁷ Partially based on their study, recent reviews of the literature have concluded that “We lack convincing evidence of algorithmic filter bubble in politics” (Guess et al., 2018).⁸ This literature is typically based on cross-sectional correlations and does exploit exogenous variation. Using experimental variation in subscriptions to outlets, this paper decomposes the mechanisms which limit exposure to counter-attitudinal news and shows that a filter bubble does exist, i.e., that conditional on subscription, Facebook is more likely to expose individuals to news matching their ideology. Furthermore, I exploit the shock generated by the experiment to show that this exposure is important since it has a strong effect on the news sites individuals visit.

This paper contributes to the literature on social media, pro-attitudinal news consumption, and polarization by testing the effect of varying the main mechanism through which social media is suspected to increase polarization: the distance between individuals’ ideology and the slant of the news they consume. Other papers on the topic have focused on the reduced-form effect of social media on polarization and have come to mixed conclusions (Allcott et al., 2019; Boxell et al., 2017; Lelkes et al., 2015). Since these papers focus on social media generally, they do not identify the causal effect of pro-attitudinal news. Indeed, a recent review of the literature argued that “it is far from clear ... that partisan news actually causes affective polarization” (Iyengar et al., 2019). To the best of my knowledge, this paper provides the first experimental evidence that counter-attitudinal news decreases affective polarization.

This study also contributes to a well-established literature on media persuasion by randomly assigning news outlets on social media in a natural setting. Survey experiments have found that individuals are persuaded by the news they consume (e.g., Coppock et al. 2018) and papers with quasi-experimental designs show that viewers are persuaded by Fox News (DellaVigna and Kaplan, 2007; Martin and Yurukoglu, 2017). In other settings, however, papers have not found that media affects political opinions (e.g., Gentzkow 2006). In a survey of the literature, Strömberg (2015) argues that it is still not clear if, and to what extent, consumers’ political behavior is affected by the news they consume. In many contexts, the “gold standard” for measuring causal effects is field experiments, since they combine the strong identification offered by lab experiments with higher external validity. However, with the notable exception of Gerber et al. (2009), there have

⁷The study focuses on posts shared by individuals’ social networks, while I focus on posts shared by outlets individuals subscribe to, which are more likely to increase segregation

⁸See also Zuiderveen Borgesius et al. (2016).

been almost no field experiments exposing participants to media outlets.⁹

Methodologically, this paper contributes to a growing literature conducting online media-related experiments (Allcott et al., 2019; Bail et al., 2018; Chen and Yang, 2018; Jo, 2018) by demonstrating how an experiment can exploit social media’s existing infrastructure to gradually distribute news to participants in a natural setting. Since the treatment occurs organically and the intervention is subtle, the treatment effects in this paper were expected to be relatively small. Therefore, I collect a sample size that is an order of magnitude larger than most other related experiments, to precisely detect the primary effects studied.

The remainder of the paper is organized as follows. In the next two sections, I provide background on Facebook and present exploratory analyses showing that social media is associated with pro-attitudinal news. Section 3 describes the experimental design, the datasets used and the empirical strategy. Section 4 analyzes the effects of the experiment on news exposure, consumption and sharing behavior and Section 5 analyzes the effects on political opinions and affective polarization. Section 6 decomposes the factors increasing exposure to pro-attitudinal news on social media. Section 7 suggests a theoretical framework explaining the effects found on opinions and attitudes. The final section concludes.

1 Background: Facebook

This study focuses on Facebook since it is the dominant social network, used by 69% of American adults. Most of these users visit Facebook several times a day and generally the platform accounts for 45% of time spent on social media.¹⁰ Despite its prominence, Facebook has been understudied, especially compared to Twitter, probably because Twitter data is more easily accessible (Guess et al., 2018; Tucker et al., 2018).

The most distinctive feature of Facebook is the news feed, where users scroll through a list of posts curated by Facebook. Posts in a user’s feed are typically shared by the user’s Facebook friends,

⁹Gerber et al. (2009) offered individuals a random subscription to the print edition of the Washington Post or Washington Times before the 2005 Virginia gubernatorial election. Their research finds no effect of the newspaper provided on knowledge, opinions or turnout, but both newspapers increased the Democratic vote share. As the authors note, the most important limitation of the study is the relatively small sample size. In addition to collecting a much larger sample and focusing on social media, this paper differs from Gerber et al. (2009) by collecting data directly measuring news exposure and consumption, allowing me to estimate the effect of exposure to news on beliefs.

Bail et al. (2018) randomize exposure to content from liberal and conservative bots on Twitter. The bots retweeted messages from various political twitter accounts, including elected officials, opinion leaders, non-profit groups and media organizations. In contrast to Bail et al. (2018) this experiment is designed to be natural as possible and thus participants were not encouraged to consume news. In addition, since I randomize exposure to actual outlets, the treatment in my experiment has a clear interpretation and is virtually equivalent to the actual treatment encountered by millions of Americans subscribing to these outlets. Finally, I collect data allowing me to analyze the chain of media effects and decompose the mechanisms leading to changes in beliefs. I further discuss the studies in Section 5.1.

¹⁰Facebook usage is based on the Pew Research Center January 2019 Core Trends Survey. https://www.pewresearch.org/wp-content/uploads/2019/04/FT_19.04.10_SocialMedia2019_toplevel_methodology.pdf

The time spent on Facebook only refers to the Facebook platform and not to other social media platform owned by Facebook, such as Instagram. Williamson, Debra Aho (2019) - US Time Spent with Social Media 2019. eMarketer.

shared by Facebook pages, such as media outlets, or are sponsored advertisements. Facebook’s algorithm determines which of the posts appear in the user’s feeds and the order of the appearance. The posts may include text, video, pictures, and links.

Facebook is a very popular source for news consumption. Approximately 43% of Americans get news on Facebook, more than twice the share getting news through other social networks (Shearer and Matsu, 2018). While this study focuses on US news consumers, understanding the effect of Facebook has global implications due to Facebook’s popularity worldwide. According to a recent report by the Reuters Institute, in 37 out of 38 middle and high-income countries surveyed, more than 20% of the population consumed news through Facebook weekly. In 25 countries at least 40% consumed news through the platform weekly (Reuters Institute, 2019).

With Facebook’s growing influence, it has faced several controversies in recent years, including an effort by the Russian-based Internet Research Agency to influence the elections, the spread of fake news during the 2016 US election cycle, and Cambridge Analytica’s attempt to assist campaigns with personally targeted ads. The concerns over each of these scandals were based on the assumption that individuals are easily persuaded by political information on social media.

2 Exploratory Analysis: Social Media News Consumption

2.1 Social Media is Associated with Pro-Attitudinal News

In this section, I present two stylized facts: I show that Facebook is associated with greater consumption of pro-attitudinal news, compared to news consumed through other means and that Facebook is associated with consumption of more extreme news.

To estimate the association between Facebook and news consumption, I rely on three datasets. First, the 2017 Comscore Web Behavior Database Panel provides a sample of the browsing behavior of approximately 93,000 US internet users. Previous studies showed that the Web Behavior Database Panel is representative of online buyers in the United States (Hortacsu et al., 2012). Comscore’s data has been used to compare offline news to online news (Gentzkow and Shapiro, 2011), but to the best of my knowledge, it has not been used to analyze news consumed through social media. For more details on how the data was processed see Appendix A.1.

Second, to determine the slant of each website, I rely on a dataset of news domains constructed by Bakshy et al. (2015). The dataset defines the slant of 500 news sites according to the self-reported ideology of Facebook users sharing articles from these websites. The dataset correlates well with other datasets measuring the slant of outlets (e.g. Gentzkow and Shapiro, 2010) and is used since it estimates the slant of a large number of outlets, at the website level. Throughout the paper, I refer to outlets on the Bakshy et al. (2015) dataset as *leading news outlets*. For more details on processing these outlets see Appendix A.2.

Third, as a proxy for individuals' ideology, I use 2008 zip code level voting data (Mummolo and Nall, 2016). I use 2008 data since the data is available at the precinct level (Ansolabehere and Rodden, 2012), and thus is relatively precise when aggregated at the zip code level. The results in this section are robust to using 2016 county-level election data and 2017-2018 donation data.

The data confirms that Facebook is an important source of news consumption. Overall, 7% of visits to leading news sites in the sample are referred to by Facebook, and the share increases to 16% among individuals who visited at least one site through Facebook. Facebook is the second most common referral source for online news sites after Google.¹¹

Figure 1 presents a binned scatter plot of news consumption by ideology. In the figure and throughout the paper, news slant is measured at the outlet level since this allows me to classify the news slant of any visit to a major news outlet and since this is the typical measure used in the literature (Gentzkow and Shapiro, 2011; Guess, 2018; Peterson et al., 2018).¹²

The figure shows a clear correlation between the consumer's ideology and the slant of the news they consume. More importantly, the slope of news consumed through Facebook (the solid blue line) is steeper than the slope for news consumed through other means (the dashed black line), indicating that news consumed through Facebook tends to better match one's ideology. To construct this figure, I calculate the mean slant of news sites visited for each individual in the sample when the sites were accessed through Facebook, i.e., the referring domain was facebook.com, and when the sites were accessed through all other means, e.g., through a search engine or by accessing the site directly. To compare a consistent sample across news referral source I only include in the sample individuals who visit news sites both through Facebook and through other means. The figure suggests that estimating segregation by comparing the news consumption of Republicans and Democrats, as is common in the literature, might mask important heterogeneity within Republicans and Democrats.¹³ Such a comparison may underestimate the association between social media and news consumption since the figure shows that individuals living in more moderate conservative and liberal zip codes consume similar news through Facebook and through other means.

Table 1 shows that the results are robust to measuring slant at the site, individual or individual by month level. Column (1) includes all domains visited and shows that an increase of 10% in the Republican vote share is associated with an increase of 0.09 standard deviations in the slant of news consumed (where a higher value is more conservative) when news is consumed through Facebook compared to other news consumed. Column (2) shows that the effect is robust to adding

¹¹Reports using different datasets confirm that Facebook is an important source of news consumption. For example, according to Parse.Ly, which collects data from a large network on online publishers, Facebook is the second most important referral source for publishers and accounted for approximately 25-30% of external traffic to digital publishers during the study period. Parse.Ly - 2018 Traffic Sources by Content Categories and Topics. For updated data on external referrals in the Parse.Ly's network see: <https://www.parse.ly/resources/data-studies/referrer-dashboard/>

¹²In a separate work in progress, I compare segregation when it is measured at the outlet and article-level.

¹³Existing measures of segregation often compare two groups of individuals, either Democrats and Republicans or liberals and conservative. See, for example, Gentzkow and Shapiro (2011); Guess (2018); Peterson et al. (2018)

individual and month fixed effects. These regressions do not take into account spillovers. Individuals could merely be switching the medium through which they consume specific domains, without being affected by social media. For example, if a conservative visits Fox News by following Facebook links, she may, as a result, consume less news by directly entering the Fox News home page. Columns (3)-(4) overcome this issue by measuring the association between the share of news consumed through Facebook and the mean slant of *all* news an individual consumed. Column (3) confirms that individuals visiting a greater share of news sites through Facebook tend to consume news better matching their ideology. Column (4) shows that the result is robust to controlling for individual and month fixed effects, and column (5) shows the result is robust to using Facebook usage as the independent variable.

While the mean slant of news consumed through Facebook is not extreme even at the most liberal and conservative zip codes,¹⁴ the share of pro-attitudinal news increases substantially when news is consumed through Facebook. When individuals living in the most conservative zip code decile visit news sites through any means besides a link in Facebook, 10% of their visits are to very conservative sites, such as *nationalreview.com*, while when they visit news sites through Facebook the figure goes up to 24%. Similarly, among individuals living in the most liberal zip code decile, 18% of news sites not visited through Facebook are very liberal (e.g. *newyorker.com*), compared to 27% of news sites visited through Facebook.

One possible implication of individuals consuming more pro-attitudinal news is that news consumption becomes more extreme. Indeed, Figure 2 presents the density of the mean slant of news consumption at the individual level and shows that the slant tends to be more extreme when news sites are visited through Facebook (e.g., it is more likely to resemble Fox News or Huntington Post and less likely to resemble USA Today or CNN). When visiting news sites through Facebook, 27% of individuals consume news that is on average more conservative than the Wall Street Journal or more liberal than the New York Times. While among all other news sites visited, only 11% of individuals consume such partisan news.

Appendix Table A.1 tests this result in a regression framework, similar to the regression presented in Table 1, and shows that the results are robust to measuring slant at the site, individual or individual by month level, to controlling for fixed effects and to different measures of Facebook usage.

2.2 Subscriptions vs. Social Network

In this section, I establish a new stylized fact: subscriptions to pages are driving the increased visits to pro-attitudinal news sites through Facebook.

To understand why social media is associated with pro-attitudinal news, I analyze data on the browsing behavior and social media feed of participants in the experiment's control group. This

¹⁴The mean slant in the most conservative zip code decile is more liberal than the Wall Street Journal and the mean slant in the most liberal zip code decile is more conservative than the New York Times.

data allows me to determine if each news site visited by the participant is visited by following a link on Facebook, and whether the link was shared by a Facebook friend or a Facebook page the participant subscribes to (the data is described in more detail in section 3.3).

Figure 3a confirms that news consumed through Facebook tends to better match the consumer's ideology. The left panel presents the news consumption of participants with a liberal ideological leaning and the right panel presents the news consumption of participants with a conservative ideological leaning.¹⁵ Throughout the paper, I define the ideological leaning of participants according to the party they identify with or lean toward. If participants do not lean toward either party, the ideological leaning is defined according to their self reported ideology, and if the ideological leaning still cannot be determined, it is defined according to the candidate the participants supported in the 2016 elections. The top row in figure 3a shows the distribution of the slant of all news sites visited, excluding news sites visited through Facebook, and the bottom row shows the distribution of news sites visited through Facebook. Both liberals and conservatives consume more extreme liberal and conservative news, respectively, through Facebook.

Next, I focus only on news consumed through Facebook and compare two mechanisms for how Facebook increases segregation in online news consumption: do individuals consume more pro-attitudinal news due to homophily in social networks (an "echo chamber" effect)? Or is there increased consumption of pro-attitudinal news due to the abundance of accessible, free media options on social media allowing consumers to personalize their news feed? The first theory is tested in the first row of Figure 3b, which presents the distribution of the slant of news sites visited through links in posts shared by Facebook friends. The second row in the figure tests the second theory by presenting the slant for news sites visited through links in posts shared by Facebook pages. Based on the control group, I estimate that 60% of visits to news sites are through Facebook pages.¹⁶

I find that sites visited through Facebook pages are driving the increased consumption of pro-attitudinal news. For example, when news sites are not visited specifically through Facebook, approximately 17% of news sites visited by conservatives are very conservative. When conservatives visit news sites through posts shared by their Facebook friends the share is similar, while when they visit news sites through posts shared by Facebook pages they subscribe to, the share of very conservative websites increases to 28%.

To conclude, the association between social media and pro-attitudinal news, shown in this section using two datasets, helps explain the public concern over social media news consumption. While the Comscore data provides a large and diverse panel which is useful in understanding news consumption habits, the analysis of the data does not provide clean identification, nor can it shed light on the implications of increased pro-attitudinal news consumption. Therefore, an experiment

¹⁵Since this sample size is smaller than Comscore's dataset, I focus on comparing liberals and conservatives to each other.

¹⁶Both posts supplied by Facebook friends and posts supplied by pages could be affected by Facebook's algorithm. The effect of the algorithm is analyzed in section 6.

generating random variation in exposure to news on social media is required. Using social media data, I find that pro-attitudinal news consumption is mostly driven by pages individuals subscribe to. The experiment described in the next section randomized those subscriptions.

3 Design and Data

3.1 Experimental Design

I recruited participants to the study in February-March 2018 using Facebook ads.¹⁷ Individuals who clicked the ads were directed to the survey landing page, where they reviewed the consent form and could begin the survey by logging in using their Facebook accounts. After logging in to the survey, and before treatment assignment, four potential liberal outlets and four potential conservative outlets were defined for each participant. The potential outlets were defined such that they did not include outlets the participant already subscribe to on Facebook, to ensure only *new* outlets would be offered to participants. Toward the end of the survey, after completing baseline questions on media habits and political beliefs, participants were randomly assigned to a liberal treatment, conservative treatment or control group, with the randomization blocked by participants' self-reported baseline ideology.¹⁸ Participants in the conservative treatment were offered to subscribe to the four potential conservative outlets, participants in the liberal treatment were offered to subscribe to the four potential liberal outlets and participants in the control group were not offered any outlets. To clarify, the intervention did not provide exclusive access to these outlets, and any individual can subscribe to these outlets on Facebook at no cost and minimum effort, regardless of the intervention. Furthermore, participants were not required to subscribe to any outlet (an encouragement design) or paid for subscriptions.

To maximize external validity, the experiment is intentionally designed to be as natural as possible. The only intervention is the initial nudge asking participants to subscribe to an outlet. Similar interventions occur naturally when individuals encounter suggestions for pages they can subscribe to, either placed by ads or by Facebook. After participants subscribed to an outlet, posts from the outlet appeared in their Facebook feeds according to Facebook's algorithm, just as they would if the participants would have subscribed to the outlet on their own. Since leading news outlets were chosen, the posts observed by the participants were also observed by other subscribers to

¹⁷The ads can be viewed in Appendix Figure A.1. For more details see Appendix A.3.1

¹⁸Randomization was blocked to increase power and to ensure balance across the main covariate expected to have prediction power when analyzing political outcomes. At the beginning of the survey, respondents were asked where they position themselves ideologically on a 7-point ideological scale from very liberal to very conservative, with an additional option for, "I haven't thought about it much." Participants were blocked in sequential order based on how they position themselves on the scale and when they answered the question. Each block is composed of three sequential participants who chose the exact same answer among the eight options in the ideological scale. The first participant in each block was randomly assigned to one of the three groups (the liberal treatment, conservative treatment or control group), the second participant was randomly assigned to one of the two remaining groups and the third participant was assigned to the remaining group.

these outlets (for example, more than 15 million people subscribe to the New York Times, one of the outlets offered). Finally, at no point were participants asked to engage with any posts or read any news content. Participants were free to make their own media choices and decide whether to read a post, click a link, share a post or unsubscribe from an outlet, just like the decisions they make regarding other posts appearing in their feed.

3.2 The Setting: Media Outlets and the News Environment

Figure 4 presents the primary outlets offered in the experiment. The primary liberal outlets are MSNBC, Slate, Huffington Post and The New York Times, and the primary conservative outlets are The Wall Street Journal, The Washington Times, Fox News and The National Review. The news outlets were chosen according to several criteria. First of all, they have a relatively clear ideological slant. Secondly, these outlets are relatively popular (Fox News and the New York Times have the second the third most subscribers among all Facebook news pages). Finally, outlets of varying quality and extremity are included to allow participants several different options when choosing whether to subscribe to an outlet and thus increase the likelihood that participants engage with at least one of the outlets offered.

If a participant already subscribed to a primary liberal outlet or a primary conservative outlet, the outlet was replaced with an alternative liberal or conservative outlet, respectively. Approximately 55% of participants did not subscribe in baseline to any of the primary conservative and liberal outlets. The effects on political beliefs are robust to including only these participants. The alternative outlets are presented in Appendix Figure A.2, and Appendix Table A.2 displays the full list of outlets, along with the number of times they were offered and the number of new subscriptions among participants who completed the endline survey.¹⁹

Figure 5 displays the most prominent men and women mentioned in posts shared by the primary outlets during the study period and shows that the outlets mostly discussed political figures. I calculate these figures by collecting all the posts shared by the outlets on Facebook during the study period, processing the text in the posts and identifying individuals using the Spacy Natural Language Processing algorithm. Unsurprisingly, President Trump is the dominant figure mentioned and is mentioned more than the next 17 individuals combined. Some of the most important political stories during the study period can be observed in the figure: President Trump's alleged affair with Stormy Daniels, Robert Mueller's investigation into the Russian government's efforts to interfere in the 2016 presidential election, Scott Pruitt's ethics scandals, the March for Our Lives Movement led by Parkland Student David Hogg, and the negotiation with North Korea's leader, Kim Jong Un. The figure also demonstrates the difference between conservative and liberal outlets. Liberal outlets focused on scandals related to the presidency and mentioned Vladimir Putin,

¹⁹To reduce the number of outlets, alternative outlets which were defined as potential outlets for fewer than 20 participants were excluded from the experiment, along with the participants for which these outlets were defined. This removed less than 0.1% of participants from the baseline sample.

Michael Cohen, Scott Pruitt, and Stormy Daniels much more often than conservative outlets.

3.3 Data Collection and Samples

The analysis of the experiment relies on three datasets: self-reported *survey data*, *Facebook data*, and *browser data*. To the best of my knowledge, this is the first study combining experimental variation with social media and browsing data.

Survey data is used to measure baseline and endline self-reported political beliefs and news consumption habits. 37,492 participants validly completed the baseline survey and 17,634 participants validly completed the endline survey.²⁰

Facebook data measures outlets participants subscribed to on Facebook and posts they shared. Participants logged in to the survey using their Facebook account, through a Facebook app created for the project.²¹ They were asked to provide separate permissions to access the list of outlets they subscribe to and the posts they share. Providing permissions was completely voluntarily and permissions could be revoked at any time and were revoked automatically approximately two months after participants logged in to the baseline or endline survey. Since data on baseline subscriptions was required to define the potential outlets for each participant, only participants who provided permissions to access their subscriptions when taking the baseline survey are included in the baseline sample.²²

Data on posts shared is used to estimate the effect on the intervention on political behavior. I collect data on all posts shared by participants that include a link or reference to a post from a news outlet. I exclude posts sharing photos, albums, music, and events.²³ Since posts shared are observable to the participant's social network or the general public, sharing posts can have a direct cost or benefit to the reputation of the participant. Analyzing shared posts provides two additional advantages: their analysis does not depend on participants completing the endline survey and there is no interaction between the experimenter and the participants when a post is shared. Approximately 92% of baseline participants provided access to the posts they shared for at least two weeks and this sub-sample (the *access posts sub-sample*) is analyzed when estimating

²⁰Participants were excluded from the study if the initial pages they subscribed to could not be observed, if they responded carelessly, if they already subscribed to too many outlets and there were no available four new liberal outlets or four new conservative outlets which could potentially be offered to them, if they took the survey a second time or if a technical error prevented some of their data from being collected. See additional details in Appendix A.3.

²¹To minimize measurement error, data from the app was collected using several methods, including a code running in the background of the baseline survey, a web service and multiple scripts that ran for the duration of the experiment.

²²Providing permission was not required to complete the survey or to be eligible for any rewards. The vast majority of participants who completed the survey provided these permissions. Participants who provided participants to access their subscriptions and revoked these subscriptions later are included in the baseline sample.

²³The remaining posts typically include a link or an embedded video. I focus on these posts since they are more likely to contain political content relevant to the experiment. However, the outlets offered to participants may also publish posts that contain only a photo and text (for example Fox News published photo with quotes related to the news without an accompanying link or video). This means that the effects I find on the number of posts shared as a result of the experiment are probably slightly lower than the actual effects.

the effect on sharing behavior. Among these participants, I observe 173,809 posts shared from leading outlets in the two weeks following the intervention. This sub-sample of participants also provided permissions to access the outlets they subscribe to for at least two weeks and thus allow me to measure whether participants unsubscribed from outlets.

To collect *browser data*, participants who completed the baseline survey using Google Chrome on a computer were asked to install a browser extension collecting data on the Facebook feed and news-related browsing behavior, in exchange for a small reward. The offer was made toward the end of the survey, but before the intervention, to ensure take-up is not affected by the treatment. The extension was created for the unique requirements of this study. To protect participants' privacy, the extension was designed to only collect the URLs of news sites visited. Approximately 2,447 of the 8,082 participants who were offered the extension, installed it. In most of the analysis of this data, I focus on a sub-sample of 1,839 participants who kept the extension installed for at least two weeks (the *extension sub-sample*).²⁴

The browser data is used to analyze news exposure, by estimating how often posts from specific outlets appeared in the participants' Facebook feeds. I attribute a post to a news outlet if it is shared by the outlet or contains links to the outlet's domains.²⁵ While the variation generated by the experiment is in subscriptions to the outlets' Facebook pages, I do not exclude news articles shared by the participant's Facebook friends, to accurately measure total exposure to news outlets on Facebook. In the two weeks following the intervention, participants in the extension sub-sample were exposed to 469,304 posts from leading outlets.

The browser data is also used to estimate the effect on the news sites participants visited. The extension can greatly reduce measurement error as other studies have suggested that individuals' self-reported media habits are more polarized than their actual media habits (Guess et al., 2017). In the two weeks following the intervention I observe 149,131 visits to leading outlets.

The browser data was only collected when participants used Facebook or browsed news sites on a computer while being signed into their Chrome account. In practice, individuals often use Facebook and browse the web on a mobile device or at work, where they may use a different browser. Therefore, all estimates of the number of times individuals were exposed to outlets in their feed or visited news sites are lower bounds for the actual intention to treat effect. In the baseline survey, participants were asked how many links to articles about government and politics they clicked on Facebook in the past 24 hours using a computer and on a mobile phone. Among

²⁴Participants were only required to keep the extension installed for two days in order to receive the reward, but most participants kept the extension installed for several weeks. In exchange for installing the extension, participants could choose between receiving a \$5 gift card, participating in a lottery with a \$200 gift card, or receiving a copy of the study results.

²⁵In order to match URLs with news outlets, I first convert over 10 million URLs to their final endpoint, allowing redirects along the way. This is required since many links in Facebook are based on URL-shortening services such as tinyurl.com. For example, I convert the url <https://tinyurl.com/y6v65aaj> to https://www.huffpost.com/entry/trump-lawyers-move-stormy-daniel-suit-to-federal-court_n_5aac5ac1e4b0c33361b0871b. For more details see Appendix A.5.

the extension sub-sample, approximately 72% of news links were clicked on a computer, so it is likely that most, but not all, data is collected for these participants.

Since the datasets are collected for a specific set of participants, the datasets define three separate sub-samples. To maximize power, throughout most of the analysis I analyze each sub-sample separately according to the outcome analyzed. When analyzing the effect on beliefs I focus on the *endline survey sub-sample*, i.e. participants who completed the baseline and endline surveys. When analyzing media outcomes, I focus on the *extension sub-sample* and the *access posts sub-sample*. Table 2 summarizes the sub-samples, the datasets used and the main outcomes.

Appendix Table A.3 presents descriptive statistics on each sub-sample. As expected, the extension sub-sample is slightly older, as the extension is only offered to participants who took the survey on a computer (in contrast to people who took the survey on their smartphone, who tend to be younger). The extension sub-sample is also more liberal and has much higher compliance rates. The samples do not differ substantially in the share of participants stating that their main source of news is social media or in the two baseline affective polarization measures.

For more details on the surveys, Facebook data and browser extension data see Appendix Sections A.3, A.4, and A.5, respectively.

3.4 Outcomes

3.4.1 Media

I measure subscriptions to outlets on Facebook, exposure to news outlets on Facebook, news sites visited and posts shared using the following quantitative outcome measures.

First, to test the direct effect of the experiment, I measure the number of times participants engaged with the *potential outlets*. For example, I measure the number of times participants observed their potential liberal and conservative outlets in their feed, and the number of time they visited the websites of their potential liberal and conservative outlets. Second, I measure the mean slant of all *leading news outlets* participants engaged with, where the slant of each outlet is based on Bakshy et al. (2015) and a higher value is associated with a more conservative slant. Third, to measure the effects of the pro- and counter-attitudinal treatments on total news consumption, I define a *congruence scale*, calculated as the mean slant of news consumption, multiplied by (-1) for liberal participants. This scale has a higher value when individuals consume more extreme content matching their ideology. Fourth, I estimate the *share of counter-attitudinal news* as an additional measure of segregation in news consumption, calculated as the share of news from counter-attitudinal outlets among all news from pro-attitudinal and counter-attitudinal outlets.

3.4.2 Opinions and Attitudes

I analyze the effect of news exposure on two primary outcomes: affective polarization and political opinions. The construction of both primary outcomes is defined in the study's pre-analysis plan along with the control covariates used in the primary regressions. The pre-analysis plan is discussed in more detail in Appendix C.

Political opinions are measured based on an index composed of twenty survey questions. The questions focus on domestic political issues and figures covered in the news during the study period, such as new tariffs, the March For Our Lives Movement, and the investigation regarding Russian interference in the elections.²⁶ Each outcome variable is defined such that a higher value is associated with a more conservative opinion and then standardized by subtracting the control group mean and dividing by the control group's standard deviation.

Affective polarization measures negative attitudes toward the opposing party. It is measured as an index composed of five outcomes. First, I use the feeling thermometer questions asking participant how they feel toward the Democrat and Republican parties (*feeling thermometer*). Second, participants are asked how well the following statement describes them on a scale from 1 to 5: "I find it difficult to see things from Democrats/Republicans point of view" (*difficult perspective*). Third, participants are asked a similar question on the following statement: "I think it is important to consider the perspective of Democrats/Republicans" (*consider perspective*). The *difficult perspective* and *consider perspective* questions are based on a political empathy index by Reit et al. (2017). Fourth, participants are asked if they think the Democrat and Republican parties have a lot (3), some (2), a few (1) or almost no good ideas (0) (*party ideas*). For each of the four previous measures, I calculate the difference between attitude toward the participant's party and attitudes toward the other party, since is a typical measure for affective polarization (Iyengar and Hahn, 2009). Fifth, to measure social-distance, participants are asked how they would feel if they had a son or daughter who married a Democrat/Republican (*marry opposing party*). This question is asked among participants who identify with a party and asks how they would feel if their son or daughter married someone who identifies with the opposing party. The answer is coded as 2 for very upset, 1 for somewhat upset and 0 for not upset at all. An advantage of using a social-distance measure is that that this measure does not correlate as strongly with other affective polarization measures, such as the feeling thermometer, and thus may be capturing an additional distinct aspect of polarization (Druckman and Levendusky, 2019). Each outcome variable is defined such that a higher value is associated with more polarization and then standardized.

I focus on affective polarization since scholars agree that this measure has been increasing over time (Gentzkow, 2016; Lelkes, 2016) and there is growing concern over its implications on political accountability, governance, sorting, professional behavior, accurate beliefs, and even product and labor markets.²⁷ While the effects of affective polarization are beyond the scope of this paper, it is

²⁶The full list of questions is presented in Appendix Figure A.8.

²⁷See Iyengar et al. (2019) for a review.

worthwhile to discuss at least two negative consequences in more detail. First, studies have shown that political behavior today is more likely to be driven by negative attitudes toward the other party instead of positive attitudes toward the voter’s party (Iyengar and Krupenkin, 2018). As a result, in recent elections, voters split their vote at record-low levels and are loyal to their party at record-high levels (Abramowitz and Webster, 2016). Consequently, elected officials may not be held accountable since they know voters on their side of the aisle will continue voting for them if they do not represent them well. Second, studies have shown the affective polarization affects economic relations. For example, in experiments workers demand a higher reservation wage when their employer is from the other party (McConnell et al., 2018), and applicants affiliated with the minority party are less likely to receive a callback when sending their resume (Gift and Gift, 2015). The increase in affective polarization has not escaped the public and in a recent survey, 85% of Americans stated that the tone and nature of political debate have become more negative over the past several years, compared to only 3% who said that the tone has become more positive (PEW, 2019).

For both the affective polarization and the political opinion outcomes, the final index is composed by taking an average of all the index components and the index is then standardized with respect to the control group so all effects are measured in standard deviations.

3.5 Balance and Attrition

Table 3 presents descriptive statistics for participants who completed the endline survey, by treatment group, and shows the sample is balanced. Since for several outcomes the analysis compares individuals by whether they were exposed to a pro-attitudinal or counter-attitudinal treatment, Appendix Table A.4 presents a balance table according to whether the treatment matched the participant’s ideology, and shows that the sample is balanced also along the re-defined treatment arms.

Similarly to other opt-in panels (Yeager et al., 2011) and other studies recruiting using Facebook ads (Allcott et al., 2019), the sample is not nationally representative. Participants tend to be more liberal than the general population, and as expected, more participants say that they get most of their news on social media (18%), compared to the national population (13%).²⁸ In contrast to the ideological composition, the gender composition of participants and their average age is very similar to the US population.

²⁸There are several likely explanations for why the sample is different from the US population. First, it is common that the samples in opt-in surveys are more liberal. Second, anecdotal evidence suggests that some conservatives who saw the ads did not want to participate in a survey conducted at Yale University. Finally, the ads automatically target people who were likely to complete the survey and not a random sample of the population. Still, the sample does not seem substantially different from samples of Mechanical Turk users, (Berinsky et al., 2012), for example. One advantage of this sample is that Facebook users are not experienced, semi-professional survey takers, in contrast to many Mechanical Turk workers. Participants were asked in the endline survey how many additional surveys they completed in the past month, the median answer is 1 and the mean answer is 7. For comparison, a 2014 study found that the median Mechanical Turk reported participating in 20 academic studies in the *week* before the question was asked (Rand et al., 2014).

Tables 3 and Appendix Table A.4 also test for differential attrition among the three endline subsamples: participants who completed the endline survey, participants who provided access to posts they shared for at least two weeks and participants who installed the extension for at least two weeks. While there are almost no differences in the attrition rates in the latter two subsamples, there is slightly greater retention of participants who completed the endline survey in the control group (48%), compared to the liberal (45%) and conservative groups (45%).²⁹ The differential attrition mostly stems from the fact that some participants in the conservative and liberal treatments did not complete the final screen of the baseline survey after they encountered the intervention, either due to a technical issue that affected a small share of participants or since they preferred not to complete the survey at that stage.³⁰

Appendix Table A.5 includes only participants who completed the endline survey and shows that despite the differential attrition, there are no substantial differences in observables between the liberal treatment, the conservative treatment, and the control group.³¹ Most importantly, there is no differential attrition in completing the endline survey between the conservative treatment and liberal treatment and no differential attrition between the pro-attitudinal treatment and counter-attitudinal treatment. In the primary endline survey outcomes I compare the two treatment arms to each other and thus the estimation does not suffer from differential attrition.

3.6 Empirical Strategy

Throughout the paper, I use two main empirical strategies. When estimating the effect of the intervention on political opinions, the slant of news participants engaged with and their engagement with the liberal and conservative outlets, I estimate the effects of the liberal and conservative treatment. When measuring the effect on polarization, it no longer makes sense to use these treatments (a conservative treatment is not expected to make participants more or less polarized than a liberal treatment), and therefore I focus on the pro-attitudinal and counter-attitudinal treatments. I also estimate the effect of these treatments on the share of counter-attitudinal news participants engaged with, the congruence scale of media they engaged with and their engagement with the pro-attitudinal and counter attitudinal outlets.

²⁹Table 3 also shows that there is a very small, but statistically significant difference between the conservative treatment and the other groups in the number of participants who provided permissions to access their posts for two weeks following the intervention (the *Access Post, Two Weeks* variable). However, this minimal difference seems to be random, since it already existed before the intervention, as can be seen by the variable *Access Post, Pre-Treat*.

³⁰Participants who did not complete the survey did not provide their email address in most cases and therefore it was more challenging to recruit them to the endline survey.

³¹Similarly, Appendix Table A.6 shows there are no substantial differences in observables according to whether participants were assigned to the pro-attitudinal treatment or counter-attitudinal treatment. In both cases, based on an F-test, I reject the hypothesis that the treatment arms are different from each other based on observables.

3.6.1 Liberal and Conservative Treatments

I estimate the effects of the liberal and conservative treatments using the following ITT regression:

$$Y_i = \beta_1 T_i^L + \beta_2 T_i^C + \alpha X_i + \varepsilon_i \quad (1)$$

where $T_i^L \in \{0,1\}$ is whether participant i is assigned to the liberal treatment, $T_i^C \in \{0,1\}$ is whether participant i is assigned to the conservative treatment, and X is a set of control variables.

As defined in the pre-analysis plan, when estimating the effect on political opinions, I focus on the difference between the liberal and conservative treatments, by testing whether $\beta_1 < \beta_2$ (i.e., the conservative treatment made participants more conservative, compared to the effect of the liberal treatment). In addition to mitigating concerns over differential attrition, comparing these treatments to each other, instead of comparing each treatment separately to the control group, leads to cleaner theoretical predictions. While the liberal outlets are clearly more liberal than the conservative outlets, it is not necessarily the case that the assigned liberal outlets are more liberal than news consumed in the control group (and similarly it is not clear if the conservative treatment exposes participants to more conservative content, compared to the control group). Finally, if individuals are persuaded by both treatments, comparing them to each other takes advantage of the experiment's design and provides more power than comparing each treatment to the control group.

To increase power, when estimating the effect on political opinions, I control for the following set of pre-registered covariates, X : self-reported ideology, party affiliation, approval of President Trump, ideological leaning, age, age squared, gender and baseline questions measuring political opinions similar to the questions used in the endline survey.³² When estimating the effect on media outcomes, I only control for baseline outcomes, when they exist. All regressions use robust standard errors unless noted otherwise.

3.6.2 Pro-Attitudinal and Counter-Attitudinal Treatments

I estimate the effects of the pro-attitudinal and counter-attitudinal treatments using the following ITT regression:

$$Y_i = \beta_1 T_i^P + \beta_2 T_i^A + \alpha X_i + \varepsilon_i \quad (2)$$

where $T^P \in \{0,1\}$ measures whether the participant was assigned to the pro-attitudinal treatment, defined as a liberal treatment assigned to a participant with a liberal ideological leaning or a conservative treatment assigned to a participant with a conservative ideological leaning. $T^A \in \{0,1\}$

³²The following baseline questions are used as controls when estimating the effect on political opinions: feeling toward President Trump on the thermometer scale, does the participant personally worry about illegal immigration, is Mueller conducting a fair investigation, has Trump attempted to obstruct the investigation into Russian interference in the election. For the definition of each control variable see Appendix B.1.

measures whether the participant was assigned to the counter-attitudinal treatment, defined as liberal treatment assigned to a participant with a conservative ideological leaning or a conservative treatment assigned to a participant with a liberal ideological leaning. As explained in section 2.2 the ideological leaning of participants is first defined according to their party, then based on self-reported ideology and finally if the participant does not lean toward any party or identify as liberal or conservative the ideological leaning is defined according to the candidate the participant supported.³³

When the dependent variable is affective polarization, $\beta_2^A < \beta_2^P$ tests whether individuals become more polarized when assigned to pro-attitudinal news, compared to counter-attitudinal news. With affective polarization as the dependent variable, I use the same set of control variables used when analyzing the effect on political opinions, with baseline measures of political opinions replaced with baseline measures of affective polarization.³⁴

3.7 Compliance

I first focus on the ITT estimates since these estimates apply to the entire sample and they do not require any additional assumption. To measure the effect of compliance, I also analyze TOT estimators by regressing the dependent variable on compliance and instrumenting compliance with the random treatment assignment. Any participant who subscribed to at least one of the outlets offered is considered a complier.³⁵ Since the intervention only offers new outlets to participants, defiers do not exist in this experiment.³⁶ Therefore, the TOT effect is simply the ITT effect scaled according to the share of compliers. The exclusion restriction is that the treatment assignment only affects outcomes through subscription to the potential outlets. It is unlikely that this assumption is violated as the only difference between the treatment arms is the subscriptions offered.

In the entire baseline sample, 56% of participants who were offered pro-attitudinal outlets complied with the treatment and subscribed to at least one outlet, compared to 45% of participants

³³Approximately 3% of participants did not self-identify as liberal or conservative, did not identify with the Republican or Democratic party and did not vote for Trump or Clinton, and are excluded from the analysis when analyzing the effect of the pro- and counter-attitudinal treatments.

All the results are robust to defining ideological leaning first by self-reported ideology, then by party and then by candidate, which how I originally defined ideological leaning in the pre-analysis plan. I prefer defining ideology first according to party to make the study comparable to other papers which tend to focus on party affiliation (Druckman and Levendusky, 2019) and since the affective polarization questions focused on Republicans and Democrats and not liberals and conservatives.

³⁴The following baseline measures are used when estimating the effect on affective polarization: baseline value of the *feeling thermometer* measure and baseline value of the *difficult perspective* measure.

³⁵Subscriptions are measured using Facebook data. Participants were also asked in the baseline survey how many pages they subscribed to. For 88% of participants, the self-reported number equals the number collected using Facebook data, suggesting data was collected properly and participants answered questions truthfully in most cases.

³⁶Defying the experiment would mean unsubscribing from an offered outlet, but participants are only offered outlets they are not already subscribed to and therefore a participant cannot defy the experiment. Since compliance is defined as subscribing to an outlet when it is offered, always-takers do not exist. When focusing on the two weeks following the intervention, an always-taker would be defined as a participant who would subscribe to the potential outlets, regardless of the intervention. In the control group only 0.6% of participants subscribed to their potential liberal outlets, and only 0.4% subscribed to their potential conservative outlets, in the two weeks following the intervention.

who were offered counter-attitudinal outlets. Appendix Table A.7 presents descriptive statistics on the compliers by treatment, and show that liberals, women, and participants who subscribe to more outlets on Facebook were generally more likely to comply with the treatments.³⁷

The difference between the share of participants subscribing to pro- and counter-attitudinal outlets is relatively small, compared to other media experiments (e.g., Iyengar and Hahn 2009) and more in line with observational studies arguing that selective exposure is not high (e.g. Guess et al. 2018). One possibility is that moderates are driving these results, however even among participants who say they are very liberal or very conservative, 44% of participants comply with the counter-attitudinal treatment.

4 Findings: Demand for News on Social Media

4.1 Individuals Willing to Engage With Counter-Attitudinal News

To test whether participants engage with the offered outlets, and specifically, whether they are willing to engage with counter-attitudinal news, Figure 6 displays the effects of the pro- and counter-attitudinal treatments on subscriptions, exposure to posts, news sites visited and posts shared from the pro- and counter-attitudinal outlets, respectively.³⁸ To keep the result comparable across media outcomes, the figure is calculated for the 1,703 participants who both kept the extension installed and provided permissions to access their posts for two weeks. Each row in the figure is estimated by regressing engagement with the four potential pro-attitudinal outlets or four potential counter-attitudinal outlets on the treatments with the control group as the reference group. For example, the first row of the first panel shows that the pro-attitudinal treatment increased the number of subscriptions to pro-attitudinal outlets by 1.94 in the two weeks following the intervention, compared to the control group, and that the effect is significant as the entire confidence interval is greater than zero.

4.1.1 Subscription to Outlets and Exposure to Posts

Panel 1 of Figure 6 shows that two weeks after the intervention, participants assigned to the counter-attitudinal treatment subscribed to 1.43 new counter-attitudinal outlets on average. This figure is similar to the initial number of subscriptions (1.52, not shown in the figure) since relatively few participants unsubscribed from outlets.³⁹

Panel 2 of Figure 6 shows that following the intervention, participants in the pro-attitudinal and counter-attitudinal treatments were exposed to more posts from the assigned outlets. While this

³⁷The number of compliers with the liberal and conservative treatments is 51% and 50%, respectively.

³⁸Appendix Figures A.4 and A.5 present the results of this subsection for the liberal and conservative treatments.

³⁹In the two weeks following the intervention, participants unsubscribed from only 6% of the outlets they subscribed to in the experiment.

effect is very strong relative to the control group, it is still a small share of the total number of post participants are exposed to in their Facebook feed.⁴⁰ To test whether participants noticed the change two months after the intervention, they were asked in the endline survey how often they saw posts from various outlets in their Facebook feed. Appendix Figure A.3 shows that participants reported seeing more posts from the outlets they were offered and participants in the counter-attitudinal treatment stated that their feed is more diverse, compared to the other treatments. This also confirms that the treatment had an effect on the sub-sample of participants who completed the endline survey and not only on participants who installed the extension.

4.1.2 Browsing Behavior - News Sites Visited

Participants could consume news from the outlets they subscribed to by simply reading the posts in their Facebook feed or by clicking links included in the posts to visit the outlets' websites directly.⁴¹ Panel 3 of Figure 6 shows that the counter-attitudinal treatment increased visits to the counter-attitudinal outlets' websites by approximately 82%, an ITT effect of 1.36 additional visits over a baseline of 1.67. The pro-attitudinal treatment increased the number of visits to pro-attitudinal outlets by 23%, an ITT effect of 2.91 additional visits over a baseline of 12.92. Due to the low number of visits to counter-attitudinal outlets in baseline, the relative effect of the counter-attitudinal treatment on counter-attitudinal outlets is greater than the effect of the pro-attitudinal treatment.

4.1.3 Sharing Behavior

Panel 4 of Figure 6 shows that participants not only consumed news from counter-attitudinal outlets when they started appearing in their feeds, they also shared the posts with their social network (albeit the effect of the pro-attitudinal treatment is not statistically significant).⁴²

To increase power, I also analyze the effect on posts shared using the entire sub-sample of participants who provided access to their posts. Figure 7 confirms that both treatments have a significant effect on the number of posts shared. Complementing previous studies focusing on Twitter (Halberstam and Knight, 2016; Gorodnichenko et al., 2018), in the control group participants are much more likely to share posts from pro-attitudinal outlets. However, the relative effect on sharing counter-attitudinal posts is stronger. The fact that participants chose to share these posts suggests that participants noticed the posts and considered them important. Sharing the posts also implies

⁴⁰Participants in the control group were exposed on average to approximately 2,162 posts in the two week following the intervention. The new posts from the offered outlets observed due to the experiment represent less than 5% of the posts in the social media feed.

⁴¹Approximately 81% of posts from outlets participants subscribed to contained links.

⁴²The wide confidence intervals stem from the large variation in the number of pro-attitudinal posts shared (the confidence interval is smaller for counter-attitudinal posts since there is little variation in how often they are shared—the vast majority of participants never share these posts)

that participants indirectly amplified the treatment effect, by expanding some of the treatment to their social network. One possibility is that participants shared posts while commenting negatively on their content. Panel 2 of Appendix Figure 7 focuses on posts which were shared with no commentary by the participants and shows that even among these posts, the counter-attitudinal treatment has a positive significant effect on the number of posts shared.

To conclude, merely offering individuals an option to subscribe to counter-attitudinal outlets causes individual to visit the outlets' websites and even share posts from the outlets.

4.2 The Social Media Feed Strong Affects Online News Consumption

The previous section demonstrates that individuals engage with the potential outlets when they appear in their feed. This suggests that search costs play a role in deciding what to consume, since the news articles participants visited, could also have been accessed without the intervention.⁴³ This raises the question of whether individuals adjust the rest of their news consumption such that their overall news diet will not change. For example, individuals randomly offered the New York Times may start consuming more articles from the outlet's website, but as a response choose to consume less news from the Washington Post, which offers a similar perspective. In this section, I focus on the conservative and liberal treatments since there are clear predictions on how these treatments would affect the mean slant of news engagement (e.g., if a conservative treatment had an effect, the feed is expected to become more conservative, while it is not clear if a pro-attitudinal treatment should make the feed more conservative or liberal).

4.2.1 Exposure to Posts

The first panel of Figure 8 shows that when participants were randomly offered liberal or conservative outlets, their feed became substantially more liberal or conservative, respectively. The change in slant is important for two reasons. First, it provides a strong first stage which is useful when analyzing the effect on political beliefs. Second, it provides an opportunity to test whether a change in the social media feed affects the slant of news sites visited or whether participants maintain a constant slant. The latter would suggest that participant re-optimize the sites they visit following an exogenous shock to their feed.

⁴³ Any participant could access the content posted by the outlets by directly going to the outlets' Facebook pages. The intervention decreased the cost of accessing these articles, by displaying them directly in the participants' Facebook feed. Another possible interpretation for the effect of the intervention is that Facebook's algorithm conveys information to the participants, e.g. the participant learned that a specific article is worth consuming since Facebook decided to supply the post with a link to the article to the consumer.

4.2.2 Browsing Behavior - News Sites Visited

I find that individuals do *not* re-optimize the slant of their news consumption. Panel 2 of Figure 8 shows that the treatment has a strong and significant effect on the slant of news sites visited by the participants. Appendix Table A.8 show that the effect is robust across various sub-samples (e.g. when excluding participants who did not complete the endline survey).

The difference between the TOT effects of the liberal and conservative treatments equals 23% of the difference between the slant of the browsing behavior of conservatives and liberals in the control group. Another way to understand the magnitude of this effect is to use the large Comscore panel to estimate the mean slant of the news individuals consume online in different states. The TOT effect of the liberal treatment would have shifted the online news consumption of an individual in Pennsylvania, a swing state, to the diet similar to an individual in New York, a blue state, while the TOT effect of the conservative treatment would have shifted her online news consumption to a news diet similar to an individual in Alabama, a red state.⁴⁴

I exploit the variation generated by the treatment to estimate the importance of the social media feed in determining the slant of news consumption. Table 4 shows that when the compliers' news feed becomes one standard deviation more conservative, the slant of the sites they visit becomes 0.31 standard deviations more conservative. The figure is calculated by instrumenting the slant of the posts observed in the Facebook feed with the treatment. In column (2) I focus only on news sites visited through Facebook (instead of all news sites visited) and find that when the feed becomes one standard deviation more conservative, the slant of sites visited through Facebook becomes 0.72 standard deviations more conservative. These regressions rely on the exclusion restriction that the treatment only has an effect on the slant of sites visited through the slant of the Facebook feed. While the treatment is only expected to affect the slant of news sites through the social media feed, the treatment could affect the feed in many ways. I am condensing the feed, a complicated object with many dimensions, to a scalar, the mean slant of news an individual was exposed to. This scalar is strongly affected by the treatment and has intuitive economic meaning, but it is possible that other changes in the feed, not captured in this measure, could affect the news sites visited. Since the calculations rely on stronger assumptions than the ITT and TOT estimates, they should be interpreted cautiously.

Figure 9a shows that the changes in media behavior decline over time but remain significant. The figure displays the difference between the effects of the liberal and conservative treatment on the mean slant of all news participants were exposed to and sites they visited, based on participants who kept the extension installed for at least six weeks. The difference between the effect of the liberal and conservative treatments in the 6th week following the intervention declines by 30% compared to the immediate effect in the first week.

⁴⁴For each individual in Comscore's database, the websites visited are matched with the leading news outlets to determine the individual's mean news consumption slant. Individuals who visited a news site only once are excluded. The slant is then calculated at the state level for all panel members in the state. The example focuses on states where there is a larger sample of at least 750 Comscore panelists who visited news sites more than once.

To test for spillovers across news outlets, i.e. whether offering an outlet to a participant increased or decreased consumption of outlets not offered in the experiment, I recalculate the effect of the treatments on the slant of news consumption and exposure excluding the eight potential experimental outlets defined for each individual. Appendix Figure A.6 shows that the mean slant of news consumption is not substantially affected by the treatments when the potential outlets are excluded, implying that the experiment did not have strong crowd-in or crowd-out effects on outlets.

4.2.3 Sharing Behavior

The third panel of Figure 8 shows that the slant of posts shared was affected by the treatment, but the effect is smaller compared to the effect on news sites visited.⁴⁵ Appendix Figure A.7 confirms the result among the entire sub-sample of participants who provided access to their posts. The effect of the liberal treatment is weaker, but the difference between the treatments is still significant. Figure 9a shows that the effect on posts is persistent, by calculating the weekly effect of the conservative treatment, compared to the liberal treatment among participants who provided access to posts for at least six weeks.

This section raises concerns regarding the power of social media companies in shaping news consumption habits. The extremely low costs of subscribing to an outlet along with the fact that individuals do not re-optimize their news consumption imply that merely suggesting several new outlets can drastically change one's news diet. Such suggestions happen all the time. They can stem from companies attempting to maximize profits by increasing user engagement, for example when platforms suggest to users outlets they may be interested in. Subscription suggestions may also originate from entities attempting to maximize political goals, whether they are NGOs purchasing ads to promote their organizations on social media or even foreign agents promoting specific pages on Facebook in attempt to influence the American electorate.⁴⁶

5 Findings: Opinions and Attitudes

5.1 Social Media News Exposure does Not Strongly Affect Political Opinions

Panel 1 of Figure 10 shows the treatment did not affect the political opinions index. The conservative treatment increased the political opinions index by 0.003 standard deviations, compared to the liberal treatment. While the point estimate has the expected sign, the effect is very small

⁴⁵The effect is significant for the conservative treatment and when comparing the treatment to each other. The difference between the control group and liberal treatment is not significant among this sub-sample.

⁴⁶For example, many ads purchased by Russian organizations in their attempt to influence the 2016 election promoted Facebook pages. See: US House of Representatives - Permanent Select Committee on Intelligence. Exposing Russia's Effort to Sow Discord Online: The Internet Research Agency and Advertisements. <https://intelligence.house.gov/social-media-content/>

economically and is not statistically significant. The upper bound for the combined liberal and conservative treatment effects, based on a 95% confidence interval, is only 0.7% of the difference in political opinions index between liberals and conservatives in the control group.

Appendix Figure A.8 shows the effect on each component in the political opinions index. The effects are economically small, and I cannot reject a null effect for any of the components.

While the previous section showed that the treatment dramatically affected the Facebook feed of participants, it is possible that no effect is detected since participants consume a small share of their news on Facebook. This explanation seems unlikely as participants who report getting most of their news on social media were affected similarly to other participants (see Appendix Section D.1). A second possibility is that the null effect masks important heterogeneity. Perhaps some participants were persuaded by the outlets they consumed, while in other cases there was a backlash effect and participants' opinions moved in the opposite direction of the treatment. For example, some conservatives may have become more conservative when exposed to liberal outlets while liberals became more liberal and as a result, the average treatment effect is close to zero. I test this hypothesis by estimating the effect of the interaction of ideology and treatment on the political opinions index and find no evidence for a backlash effect. Appendix Figure A.9 shows that liberals did not become significantly more liberal as a result of the conservative treatment and conservatives did not become more conservative as a result of the liberal treatment.

Finally, to verify that the magnitude of the estimate is economically small I use the treatment as an instrument for the slant of participants' Facebook feed to estimate how the feed's slant affects opinions. I find that if the Facebook feed of a liberal became similar to the Facebook feed of a conservative over a two month period (or vice versa), the opinions of the liberal would move only 3% in the direction of the opinions of the conservative (see Appendix Table A.9), and I can reject an effect larger than 7% .

Interestingly, the results differ from a recent study by Bail et al. (2018), which exposes individuals to different views on Twitter and finds evidence for a backlash effect. Two major differences between the studies can explain the differing results. First, Bail et al. expose individuals to different *views* on Twitter and not only to news outlets. Individuals plausibly became more upset when they are exposed to the twitter feed of opposing elected officials and opinion leaders, compared to counter-attitudinal news outlets. Second, Bail et al. provided participants with financial incentives to continuously follow the new information they were exposed to. While financial incentives provide a stronger treatment effect, they may also encourage individuals to continue consuming news which increases their partisan hostility, and that they would not have consumed without the incentive.

5.2 Exposure to Counter-Attitudinal News Increases Affective Polarization

In contrast to the null effect of the liberal and conservative treatments on political opinions, Panel 2 of Figure 10 shows that the counter-attitudinal treatment decreased affective polarization by 0.03 standard deviations, compared to the pro-attitudinal treatment. The TOT effect is approximately 0.06 standard deviations. This section implies that exposure to pro versus counter-attitudinal news affects partisan hostility and the concerns over increased exposure to pro-attitudinal news are not misguided.

Appendix Table A.10 shows that the result is robust to not controlling for covariates, and Appendix Table A.11 show it is robust to excluding participants who already subscribed to at least one of the primary outlets before the intervention. Appendix Table A.12 show that an effect is detected even when focusing only on the sub-sample of participants who completed the endline survey and installed the extension. The effect is stronger among this group, partially due to higher compliance rates.

Figure 11 presents the results of regressions estimating the effect for each measure in the affective polarization index separately. The effect is especially pronounced for the question asking participants how difficult they find it to see things from each party's point of view and the effect is weakest when participants are asked if parties have good ideas. While the experiment is underpowered to detect a statistically significant effect for each coefficient separately (and that is one of the reasons the pre-analysis plan stated that an index will be analyzed), in all cases the pro-attitudinal treatment is associated with a more polarized outcome and the coefficients are similar in magnitude to the point estimate of the index measure.

Appendix D.1 shows that the effects are relatively homogeneous across liberals and conservatives, participants who follow the news more and less often, older and younger individuals and generally most characteristics I test for. Appendix Figure A.10 shows that the treatment arms affected the polarization index in a consistent fashion: the counter-attitudinal treatment decreased polarization and the pro-attitudinal treatment increased it, albeit only the former effect is statistically significant.

Based on eight pre-registered survey questions I test if the effect on polarization could be explained by changes in the knowledge of participants. I do not find evidence for strong effects on knowledge. Appendix D.2 discusses this estimation in detail.

In the rest of this section, I interpret the magnitudes of the effect using three approaches. First, I compare the effect of the intervention to benchmarks in the control group and outside the experiment. Second, I use the browser data to estimate the effect of a change in exposure to pro- and counter-attitudinal news on affective polarization. Third, I conduct two back-of-the-envelope calculations to estimate how affective polarization would have changed if Facebook had more balanced news exposure. All of the estimates are based on the effects over a two-month period. It is possible that with longer exposure to different news, the effects would have been stronger.

To compare the results to existing benchmarks, I focus on the feeling thermometer question, which is asked regularly in the American National Election Surveys. The ITT effect of the counter-attitudinal treatment decreases the difference between the feeling toward the participant's party and the opposing party by 0.58 degrees (on a 0-100 scale), and the TOT effect decreased the gap by 0.98 degrees. For comparison, in the past 20 years, the feeling thermometer measure (the difference between how individuals view their own party and the opposing party) increased by 3.83 degrees. An additional point of comparison is a recent experiment which found that disconnecting from Facebook decreases the feeling thermometer measure by 2.09 degrees (Allcott et al., 2019). Hence, one way to interpret these results is that approximately half of the depolarizing effect of disconnecting from Facebook can be achieved by replacing 1-4 subscriptions to pro-attitudinal outlets with subscriptions to counter-attitudinal outlets.⁴⁷

To estimate the effect of a change in exposure to pro or counter-attitudinal news, I focus on participants who both installed the browser extension and completed the endline survey (i.e., I analyze the overlap between the extension and the endline sub-samples). I use two summary statistics for exposure to pro- and counter-attitudinal news: the share of counter-attitudinal news among all pro and counter attitudinal news and the congruence scale, where a higher value is associated with a more conservative slant of news exposure for conservative participants and a more liberal slant for liberal participants. I calculate these statistics based on all posts observed between the baseline and endline survey, for participants who observed at least two pro or counter-attitudinal posts. Each statistic is instrumented with the treatment to measure its effect on affective polarization. The IV regressions determining the effect of news exposure rely on the exclusion restriction that the treatment only has an effect on affective polarization through these statistics. Similarly to the discussion in section 4.2, it is likely that the treatment affected beliefs only through changes to the social media feed, but it is possible that other changes in the feed, not captured in these summary statistics affected polarization, and therefore the calculation relying on these statistics should be interpreted cautiously.

I find that an increase of one standard deviation in the share of exposure to counter-attitudinal news decreases affective polarization by 0.14 standard deviations. Similarly, an increase of one standard deviation in the congruence scale decreases affective polarization by 0.11 standard deviations. One challenge in studying affective polarization is determining whether the correlation between pro- and counter-attitudinal news exposure and affective polarization is due to selection, i.e., individuals with more negative views of the opposing party select into more pro-attitudinal news exposure, or a causal effect, i.e., pro-attitudinal news makes people more polarized. The effects of news exposure on affective polarization are approximately 28%-35% of the coefficients obtained using a cross-sectional regression among the control group, suggesting that the correla-

⁴⁷This interpretation ignores the small differences between the settings of the studies and the samples. I estimate an effect over two months in the spring of 2018, while Allcott et al. (2019) conduct the study over a one-month period in the fall of 2018. Furthermore, while both samples were recruited using Facebook ads, the sample compositions could still differ, for example since Allcott et al. (2019) screen respondents who report using Facebook for less than 15 minutes per day, or who are not willing to deactivate Facebook for 24 hours.

tion is both due to a causal effect and selection (see Table 5).

Finally, I use two back-of-the-envelope calculations to estimate how affective polarization would have changed if Facebook had more balanced news exposure. In Appendix Table A.13, I find that if Facebook had an equal share of pro-attitudinal and counter-attitudinal news, affective polarization would decrease by 0.19 standard deviations and the feeling thermometer measure would decrease by 3.76 degrees. For this calculation, I rely on the difference between the share of exposure to counter-attitudinal news in the control group, 17%, and an exposure of 50%. I then estimate the effect of this difference on affective polarization based on the IV regressions described above. The estimation does not rely on out-of-sample predictions as the share of counter-attitudinal news was greater than 50% for a non-negligible share of participants in the counter-attitudinal treatment.

However, perhaps having a balanced news feed is not a realistic counterfactual since most individuals do not consume balanced news, regardless of social media. Therefore, in a second back-of-the-envelope calculation, I estimate how affective polarization would change if news consumed through Facebook had a similar congruence scale to news consumed through other means. In Appendix Table A.14, I find that affective polarization would decrease by 0.06 standard deviations and the feeling thermometer outcome would decrease by 1.11 degrees.⁴⁸ These back-of-the-envelope calculations should be interpreted carefully since they do not take into account general equilibrium effects.⁴⁹ Nevertheless, they suggest that the Facebook feed can play an important role in amplifying or mitigating polarization.

6 Findings: Why is Social Media Associated with Pro-Attitudinal News?

Since the previous section shows that exposure to pro-attitudinal news affects partisan hostility, it is important to understand what influences the news individuals are exposed to on social media. This section decomposes the segregation in exposure to posts shared by pro- and counter-attitudinal outlets offered in the experiment. As shown in Panel 2 of Figure 6 the pro-attitudinal treatment increased exposure to pro-attitudinal posts by approximately twice as much as the effect of the counter-attitudinal treatment on counter-attitudinal posts.

I decompose the increased exposure to pro-attitudinal posts into three main forces: Participants

⁴⁸The result is based on the following calculation: First, I find the difference in the control group between the congruence scale of news sites visited through Facebook and the congruence scale of all other news sites. Then I calculate the effect of a change in the congruence scale of the Facebook feed on the congruence scale of news sites visited. This is calculated by regressing the congruence scale of sites visited on the congruence scale of the Facebook feed, with the feed instrument by the treatment. Using these two numbers I estimate by how much the congruence scale of participants' Facebook feed would have to decrease in order for the participant to consume news through Facebook with the same congruence scale as other news consumed. Finally, I estimate the effect of such a decrease in the congruence scale on affective polarization.

⁴⁹For example, it is likely that if Facebook drastically changed its feed, individuals would use other social networks instead. Some of this effect may be captured in the counterfactuals since participants in the counter-attitudinal treatment did use Facebook less often (as discussed in Section ??). However, with network effects, it is likely that the decrease in Facebook use would be much greater if such a policy would be implemented across the platform.

are less likely to subscribe to counter-attitudinal news outlets (“selective exposure”); Facebook’s algorithm supplies fewer posts from pro-attitudinal outlets, conditional on participants subscribing to them (the “filter bubble”); and participants use Facebook less often when offered counter-attitudinal outlets. The decomposition exercise is based on the following framework:

$$E_{ij} = S_{ij}F_{ij}U_i$$

where E_{ij} is the number of posts from outlet j individual i was exposed to. Exposure is a product of whether individual i subscribes to an outlet j (S_{ij}), the share of posts supplied from the outlet among all posts the individual observes (F_{ij}) and the total number of posts individual i is exposed to in Facebook (U_i). I decompose the difference in exposure using the following formula:

$$\Delta E = \underbrace{S_{\Delta} * F_C * U_C}_{\text{Selective Exposure}} + \underbrace{S_C * F_{\Delta} * U_C}_{\text{Filter Bubble}} + \underbrace{S_C * F_C * U_{\Delta}}_{\text{Time on Platform}} + \underbrace{S_{\Delta} * F_{\Delta} * U_C + S_C * F_{\Delta} * U_{\Delta} + S_{\Delta} * F_{\Delta} * U_{\Delta} + S_{\Delta} * F_{\Delta} * U_C}_{\text{Combinations}} \quad (3)$$

where for each variable, the $_C$ subscript notates the value for the counter-attitudinal treatment and the $_{\Delta}$ subscript notates the difference between the pro- and counter-attitudinal treatments.

- ΔE is the difference between exposure to posts from the offered pro-attitudinal outlets in the pro-attitudinal treatment and exposure to posts from the offered counter-attitudinal outlets in the counter-attitudinal treatment.
- S is estimated by pooling all potential outlets and participants and regressing subscription to an outlet on the full interaction of whether an outlet was offered to the participant and whether the outlet is pro-attitudinal (Appendix Table A.15, column 1).
- F is estimated by pooling all potential outlets and participants and regressing the share of posts supplied from an outlet on the full interaction of whether the participants subscribed to the outlet and whether the outlet is pro-attitudinal, with subscription to an outlet instrumented with whether the outlet was randomly offered (Appendix Table A.15, column 2).⁵⁰
- U is estimated by regressing the total number of posts participants were exposed to on whether they were assigned to the pro or counter-attitudinal treatment (Appendix Table A.15, column 3).

In Equation 3, *Selective exposure* measures the additional posts from counter-attitudinal outlets participants in the counter-attitudinal treatment would have been exposed to if they would have subscribed to the same number of outlets as participants in pro-attitudinal treatment. The *filter*

⁵⁰ F is estimated using two IV estimators, and thus its causal interpretation relies on the assumption that there is no essential heterogeneity (Heckman et al., 2006). Otherwise, the difference between exposure in the pro-attitudinal and counter-attitudinal treatments might be due to treatment heterogeneity and selection into compliance, and not due to different treatment effects. In a future robustness section, I plan on re-weighting the IV estimators according to the baseline covariates of compliers (Angrist and Fernandez-Val, 2013; Aronow and Carnegie, 2013) in order to estimate the ATE for never-takers and for the entire sample, and to test whether this affects the decomposition results.

bubble measures the additional posts subscribers to counter-attitudinal outlets would have been exposed to if Facebook’s algorithm would have supplied them with the same share of posts from these outlets, as the share supplied when subscribing to pro-attitudinal outlets. The *time on platforms* measures the additional posts participants from the counter-attitudinal treatment would have been exposed to if they would have used Facebook as much as participants assigned to the pro-attitudinal treatment.

I find evidence for all three effects: participants prefer to subscribe to pro-attitudinal news outlets, participants are supplied with a smaller share of posts from counter-attitudinal outlets, conditional on subscription, and participants decrease their Facebook usage after they are offered to subscribe to counter-attitudinal outlets. The last effect is only significant at the 10% level and I interpret it as suggestive evidence that participants use social media less often when they are exposed to more news they disagree with.

Figure 12 shows that the strongest force driving exposure to pro-attitudinal news in the experiment is the supply of posts by the platform. This provides clear evidence that a filter bubble exists in social media. Even when individuals are willing to subscribe to outlets with a different point of view, Facebook’s algorithm is less likely to show them content from those outlets. The result also suggests why personalization is leading to segregation online—when consumers are exposed to more counter-attitudinal news, they seem to decrease their Facebook usage and therefore platforms may have an incentive to provide individuals with news that matches their opinion, in order to maximize engagement. This result raises the question of whether the personalization of content also occurs within an outlet. A platform could supply conservatives with the more conservative articles posted by an outlet and supply liberals with the more liberal articles posted by the same outlet. I find no evidence for personalization within an outlet.⁵¹

This section does *not* suggest that Facebook’s algorithm intentionally increases segregation by targeting posts according to whether they share the consumer’s beliefs, or that the interaction of the slant of a post and ideology of a user has a causal effect on whether a post is supplied by the algorithm. It is more likely that the platform learns what the consumer may be interested in based on the consumer’s behavior, her social network and other pages she is subscribed to. By combining these factors, the platform’s algorithm may determine that posts from specific outlets

⁵¹I focus on the subset of articles that were shared by at least one Member of Congress in 2018 and define the slant of an article according to the mean DW-Nominate score of Congress Members who shared the article. Using this measure, I find that in general conservative participants are exposed to more conservative articles on Facebook, even when controlling for the outlet. This is not surprising as a conservative is likely to have more conservative friends, who are likely to share more conservative articles within an outlet. However, when I focus only on posts from the eight potential outlets defined for each participant, I do not find any correlation between the slant of the articles and consumers’ ideologies (see Appendix Table A.16). This suggests that Facebook’s algorithm does not lead to conservatives being supplied with more conservative articles, *within* an outlet. It also suggests that conservatives and liberals were exposed to similar content from the outlets they subscribed to in the intervention, conditional on a post from the outlet appearing in their feed. Data collected in December 2018.

The list of the Facebook pages of Members of Congress is based on the Congress Members project (<https://github.com/unitedstates/congress-legislators>). The list was used to collect all posts shared by Members of Congress in 2018. The list of tweets shared by Members of Congress is taken from the Tweets of Congress project (<https://github.com/alexlitel/congresstweets>). Downloaded on December 17, 2018.

are less likely to interest a user, and these outlets tend to be counter-attitudinal. Still, the effect of personalization on news exposure is an important departure from how news was supplied and consumed in the past. While I focus on Facebook, the logic probably applies to other platforms that personalize content as well. For example, since 2016 Twitter has been ranking tweets according to how interesting and engaging they would be for a specific user and the highest-scoring tweets are shown at the top of a user’s timeline. If Twitter’s ranking algorithm is similar to Facebook, this may increase exposure to pro-attitudinal news.⁵² Furthermore, major news outlets have also started to personalize their websites and the articles they suggest to their customers.⁵³

7 Interpretation

How should we interpret the fact that individuals attitudes toward parties change, while their political opinions remain relatively stable? In this section, I suggest a framework explaining the experiment’s results. Since the results are consistent with a secular trend of a sharp increase in affective polarization over the last couple of decades, while political opinions have not changed dramatically (Gentzkow, 2016; Lelkes, 2016; Mason, 2015), this framework can also explain how increased exposure to pro-attitudinal news over time could have increased affective polarization.

Consider the following model of media persuasion, based on DellaVigna and Kaplan (2007). There is some state of the world θ . Consumer i has a prior on the state of the world $\theta_i \sim (\theta_i^0, \frac{1}{h_i})$, where θ_i^0 is the consumer’s initial belief and h_i is the precision of the belief.

I extend the model by assuming that a consumer’s political opinion is a weighted average of K beliefs:

$$\gamma_i = \sum_{k \in \{1..K\}} w_{ik} \theta_{ik}$$

where $0 \leq w_{ik} \leq 1$ is the weight consumer i places on belief k when determining her political opinion. A weight can be thought of as the priority the consumer places on the specific belief. For example, a consumer’s support for a climate bill to price carbon will probably depend on whether she believes the bill will mitigate emissions and whether she believes it will increase electricity prices. A liberal might put more weight on the effect of the bill on emissions while a conservative may care more about prices.⁵⁴

⁵²Factors taken into account when determining the relevancy of tweets include the tweet’s author and the user’s past relationship with the author, therefore it is likely that tweets from pro-attitudinal account will receive a higher ranking. For more details see: Using Deep Learning at Scale in Twitter’s Timelines. https://blog.twitter.com/engineering/en_us/topics/insights/2017/using-deep-learning-at-scale-in-twitters-timelines.html

⁵³In 2017 the New York Times announced that it will tailor its homepage to the interests of individual readers. See: The New York Times. A ‘Community’ of One: The Times Gets Tailored. March 18, 2017.

⁵⁴According to a 2019 PEW survey, 74% of Democrats, including respondents leaning toward the Democratic party, state that the environment should be a top priority for President Trump and Congress in 2019, compared to only 31% of Republicans or those leaning toward the Republican party. On the other hand, 79% of Republicans say the economy should be a top priority, compared to 64% of Democrats. Pew (2019) - Public’s 2019 Priorities: Economy, Health Care,

Outlet j receives a signal on the state of the world $s \sim N(\theta^*, \frac{1}{h_s})$, where s is the signal, θ^* is the true state of the world and h_s is the precision of the signal received. While all outlets receive the same signal, each outlet reports the signal with a bias according to the outlet's ideological slant: $r_j = s + b_j$ where a larger b_j is a more conservative bias. The ideological slant of outlets can be explained by owner incentives (Anderson and McLaren, 2012), or by an attempt to maximize market share (Gentzkow and Shapiro, 2006).

Since consumers know that the outlets are biased, they do not take the reports at face value, but instead, interpret them as $f(r_j, b_j)$. The consumer's posterior is the weighted average of her prior and the adjusted report: $\theta_i^1 \sim N(\frac{h_i \theta_i^0 + h_s f(r_j, b_j)}{h_i + h_s}, \frac{1}{h_i + h_s})$.

Applying the framework to the experiment leads to a clean prediction. If a consumer is rational and completely takes into account the slant of the outlet $f(r_j, b_j) = s$ and it should not matter if she is assigned to the liberal or conservative treatment when updating her priors. Furthermore, assuming individuals in the control group are still exposed to news, they should also update their beliefs similarly. If the consumer is naive, she does take into account the slant of the outlet and should become more conservative when exposed to conservative outlets, compared to the effect of exposure to liberal outlets.

In contrast to political opinions and persuasion, most of the literature on affective polarization is new and there is no overarching framework for the topic. A straightforward way to model the attitude of individual i toward party p is to define it as a function of the distance between the political opinion of the party (γ_p) and a benchmark for the "correct" opinion according to individual i , $\hat{\gamma}_{ip} = \phi(\theta_{i1}, \dots, \theta_{ik}, w_{i1}, \dots, w_{ik}, \theta_{p1}, \dots, \theta_{pk}, w_{p1}, \dots, w_{pk})$, where $\hat{\gamma}_{ip}$ is the opinion individual i thinks party p should hold. Assume affective polarization is a linear function of the distance between the opinion of the party and the benchmark: $g(\gamma_p - \hat{\gamma}_{ip})$. I consider two functions ϕ determining the benchmark opinion.

- **Affective polarization due to political distance:** $\hat{\gamma}_{ip} = \gamma_i = \sum_k w_{ik} \theta_{ik}$

Consumers who only care about political opinions will use their own opinion as the benchmark for the correct opinion and determine their attitude toward a party based on the difference between their political opinion and the party's political opinion. When a political opinion changes from γ_i^0 to γ_i^1 , the following change is expected in affective polarization

Education and Security All Near Top of List.

As a clarifying example for the framework, I intentionally focus on a very general topic—support for climate change policy. Some of the questions forming the political opinions index are on more specific topics, but the same logic holds. For example, participants' favorability of the March for Our Lives Movement could depend on their beliefs on whether gun violence is a serious problem, whether banning certain weapons will help deal with the problem, on whether gun owners may not be able to purchase their preferred guns if the movement accomplishes its goals, and whether the leaders of the movement attack gun-right supporters. Consumers will probably place different weights on different beliefs. For example, a conservative might place a higher weight on how she believes the movement will effect gun-owners, while a liberal may place lower weights on those beliefs.

$(\Delta A_i)^{55}$

$$\Delta A_i = g(\gamma_i^1 - \gamma_p) - g(\gamma_i^0 - \gamma_p) = g(\gamma_i^1 - \gamma_i^0) = g\left(\sum_k w_{ik}(\theta_{ik}^1 - \theta_{ik}^0)\right) \quad (4)$$

This theory predicts that an update in the consumer's beliefs should only affect attitudes toward a party through its effect on the consumer's political opinions. Returning to the climate bill example, a consumer would determine her attitude toward a political party based on the distance between her support for the climate bill and the party's support for the bill. If a consumer is randomly exposed to liberal outlets and as a result her support for a climate bill increases, her attitude toward the Democratic party should become more positive and her attitudes toward the Republican party should become more negative. This theory is not consistent with the experiment since attitudes changed without a corresponding change in political opinions.

- **Affective polarization due to unreasonable opinions:** $\hat{\gamma} = \sum_k w_{pk}\theta_{ik}$

Alternatively, consumers may judge whether the political opinion of a party is reasonable according to the party's own weights. Hence, the benchmark opinion is the opinion the party would hold based on the consumer's beliefs regarding the state of the world, weighted by the weights party p places on those beliefs. According to this theory, individuals develop negative attitudes toward a party when even according to the party's weights, the party should change its political opinion. In other words, affective polarization increases when consumers cannot rationalize the parties political opinions and perceive that the party is not adhering to its own values. The change in affective polarization following updated beliefs is:

$$\Delta A_i = g\left(\sum_k w_{pk}\theta_{ik}^1 - \gamma_p\right) - g\left(\sum_k w_{pk}\theta_{ik}^0 - \gamma_p\right) = g\left(\sum_k w_{pk}\theta_{ik}^1 - \sum_k w_{pk}\theta_{ik}^0\right) = g\left(\sum_k w_{pk}(\theta_{ik}^1 - \theta_{ik}^0)\right) \quad (5)$$

If the consumer and the party place the same weight on beliefs, there is no difference between the two theories. However, if there are heterogeneous weights, political opinions and affective polarization may be differentially affected. In the climate bill example, a liberal who believes the climate bill will mitigate emissions and *decrease* consumer prices will support the bill. The consumer will have a negative attitude toward a party opposing the bill since even if the party does not place a high weight on decreasing emissions, it should support the bill. If the liberal is exposed to conservative outlets and learns that the bill is more likely to increase prices, she may still support the bill since she places higher weights on mitigating emissions but will develop a less negative attitude toward a party that places a high weight on consumer prices and thus opposes the bill. In other words, the liberal consumer will better understand the rationale behind the argument objecting to the bill even if she does not agree with the importance of the argument.

⁵⁵I am assuming without loss of generality that $\gamma_i^p < \gamma_i^0$ and $\gamma_i^p < \gamma_i^1$

This theory is consistent with the results of the experiment if the consumers updated beliefs on which they place relatively low weights, but the parties place higher weights.⁵⁶ As a result, consumers' political opinions would not change substantially, but attitudes toward parties would change.⁵⁷

To further test the theories, I analyze the effect of the experiment on attitudes toward parties. Affective polarization is usually measured as the difference between individuals' attitudes toward their own party and their attitudes toward the opposing party, but we can focus on each party separately. If media outlets are delegates of their consumers (Gentkow et al., 2015), we would expect pro-attitudinal outlets to cover more issues for which $w_{OWN} > w_{OPPOSING}$ and counter-attitudinal outlets to cover more issues for which $w_{OPPOSING} > w_{OWN}$, where w_{OWN} are the weights placed by the individual's party and $w_{OPPOSING}$ are the weights placed by the opposing party.

The two theories differ in their predictions regarding the effect of the treatments on attitudes toward each party. If affective polarization is merely a function of political distance, attitudes toward parties will mostly be affected when consumer i updates beliefs on which she places higher weights (see equation 4). Therefore, attitudes toward both parties should be strongly affected by exposure to pro-attitudinal outlets, that are more likely to cover beliefs on which i places high weights. On the contrary, if affective polarization is a function of unreasonable opinions, attitudes toward party p will be affected more by beliefs on which p places higher weights (see equation 5). Therefore, pro-attitudinal outlets are expected to have a greater effect on one's attitudes toward their own party, while counter-attitudinal outlets will have a greater effect on attitudes toward the opposing party. Table 6 shows that attitudes toward the opposing party are indeed more likely to be affected by exposure to counter-attitudinal outlets, supporting the theory that affective polarization is due to perceived unreasonable opinions. This result also contradicts a third hypothesis (not formally presented) which argues that affective polarization is increasing because of negative coverage of pro-attitudinal outlets focusing on the opposing party (Iyengar et al., 2019).

To conclude, there is still limited evidence on whether exposure to pro- and counter-attitudinal

⁵⁶There are two explanation for why consumers would updates beliefs heterogeneously. First of all, it is possible that consumers are naive when updating some beliefs and sophisticated when updating other beliefs, either since consumers are motivated to be more careful when updating beliefs they cares about, or since they have more expertise in these topics allowing them to identify media bias. Alternatively, it is possible that consumers are generally naive when updating all beliefs, but they have much stronger priors regarding beliefs on which they place higher weights when determining their political opinions.

Alternatively, it is also possible that consumers do not update their beliefs regarding the state of the world, but rather learn the weights the parties place on issues. Learning the weights allows the consumers to rationalize the opinions of the opposing party and thus is expected to decrease polarization.

⁵⁷As long as weights are positive, this theory still predicts that there will be come change in political opinion. However, it is reasonable that some consumers place zero weights on some beliefs. For example, people who do not believe climate change is happening may place a zero weight on whether a climate bill aimed decreases greenhouse gas emissions. Similarly, some liberals may not care about whether the March for Our Lives movement has a negative effect on the welfare of gun owners. More importantly, the logic behind the theory does not rely on a knife edge case, and still holds if consumers place a positive but very small weight on beliefs. In that case we would expect political opinions to be slightly affected, but the effect would be small, and thus may not be detected by the experiment (indeed the point estimate of the effect of the treatments on political opinions is positive, but economically very small).

news has an effect on affective polarization, let alone an understanding of the channel explaining this effect. This section provides evidence ruling out several theories: it is unlikely that affective polarization simply increases due to a growing difference in political opinions or that affective polarization is mostly explained by increased negative coverage in the news media. I present a parsimonious theory that is consistent with the results: consumers determine their attitudes toward a party based on the distance between the party's opinions and the opinion the party should hold according to the consumers' beliefs and the party's weights. The theory predicts that affective polarization will decrease when individuals are more likely to believe the arguments supporting the other side's point of view, even if that will not change their ultimate political opinions due to differences in priorities (weights). While I provide evidence supporting the theory, there could be other explanations for the change in affective polarization,⁵⁸ and more research is needed to pinpoint the precise mechanisms explaining how affective polarization evolves.

8 Conclusions

Consumption of news through social media is increasing, but the effect of social media on public opinion remains controversial. This paper sheds light on how social media affects news consumption, political opinions, and affective polarization.

The study shows that individuals are willing to be exposed to engage with new viewpoints through social media. Participants in the experiment do not only subscribe to counter-attitudinal news outlets, but they also consume and share news from those outlets. However, the algorithms determining news exposure mitigate potential exposure to counter-attitudinal news. This "filter bubble" effect did not exist until recently and may have stronger impacts in the future, with the development of more sophisticated machine learning algorithms customizing news exposure. This is an important development since I find that the social media feed has a large influence on online news consumption habits.

The study suggests that a more nuanced view is needed regarding the effect of media on political beliefs. On the one hand, exposure to pro-attitudinal news increases affective polarization, compared to exposure to counter-attitudinal news. Therefore, there is room for concern that exposure

⁵⁸For example, Mason (2015) explains that partisan bias may increase without changes in position extremity as a result of stronger partisan identity. In other words, the intervention may have amplified or mitigated tribalism, which can increase affective polarization, without having a strong effect on political opinions. In Appendix Figure A.3 I find that the conservative treatment did not increase identification with the Republican Party. The liberal treatment is associated with a slight increase of identification with the democratic party, however the point estimate is small, it is precisely estimated and it is not statistically significant.

An alternative explanation is that consumers learned through exposure to counter-attitudinal outlets that the other party is not as extreme as they originally thought, and thus developed less negative attitudes toward the party. In Appendix Figure A.3 I find that pro-attitudinal party does not decrease the distance between participants baseline ideology and their party's perceived ideology. Similarly, the counter-attitudinal does not decrease the distance between the participant's ideology and the other party's perceived ideology. The point estimate is not significant and moves in the opposite direction.

to news on social media is leading to more negative attitudes across parties. On the other hand, it seems that individuals are not so easily persuaded by the political leaning of their news exposure. This result should give us pause before claiming that elections are clearly determined by manipulation of opinions on social media. The fact that the results are in line with long term trends in affective polarization and political opinions suggests that a segregated news environment may partially explain the increase in affective polarization over the past several decades.

Similarly to other randomized control trials, one should be careful when extrapolating the results of the study. The experiment took place in the first half of 2018. In this period, Facebook was often criticized for news-related content appearing on the platform, and the company seemed to have responded by decreasing the exposure to news outlets on the social media feed. It is possible that in a different period, with greater exposure to news, the effects would have been larger. Similarly, Trump's presidency is exceptional in the stability of the president's approval ratings.⁵⁹ If other opinions were relatively stable throughout the period as well, the minimal effect found on political opinion may be explained by the period when the survey took place. Finally, the experiment compared outlets with an ideological slant, however, a large share of news consumption is from moderate outlets. It would be interesting in a future study to randomize subscriptions to outlets that are not liberal or conservative leaning, such as USA Today, and measure their effects on beliefs. While acknowledging these limitations, the experiment has high external validity when it comes to analyzing actual behavior on Facebook in 2018, as supply and consumption of news occurred just as they do when individuals subscribe to any other outlet on Facebook.

Finally, this study has important policy implications. Scholars have suggested that new tools are required to expose individuals to more counter-attitudinal news. Suggestions include making algorithms more transparent, allowing users an option to diversify their feed and adding modifying algorithms such that they encourage serendipitous encounters (Sunstein, 2017). The experiment described in this paper essentially measures the effect of one such intervention and shows that a simple nudge can be effective since individuals are willing to engage with other viewpoints. Social media platforms can adopt similar policies and have recently started rolling out features which could potentially diversify the users' feeds.⁶⁰ Such policies could have positive externalities. For example, if social media companies were to show users leading articles from outlets that usually do not appear in their feed, individuals might find it easier to understand other opinions and develop less negative attitudes toward the opposing parties.

⁵⁹For example: Dann, Carrie and Murray, Mark - NBC/WSJ poll: Trump Approval 'Remarkably Stable' After a Stormy Week of Bad News. NBC News. August 26, 2018; Yokley, Eli - New Poll Suggests Voters Have Made Up Their Minds on What Trump Is Like. Morning Consult. July 26, 2018.

⁶⁰For example, in 2017 Facebook implemented a feature which showed users articles from additional outlets related to a post in their feed. Facebook Newsroom - New Test With Related Articles. April 25, 2017. <https://newsroom.fb.com/news/2017/04/news-feed-fyi-new-test-with-related-articles/>.

In August 2018 Twitter announced that it will allow users to follow topics instead of specific accounts, which may also expose users to more diverse news sources. Newton, Casey - Twitter tests letting users follow topics in the same way they follow accounts. The Verge. August 13, 2019. <https://www.theverge.com/2019/8/13/20804476/twitter-interests-follow-topics-feature-accounts-timeline>

To conclude, while social media may be increasing affective polarization through the increased consumption of pro-attitudinal news, it also has the potential to mitigate these effects.

References

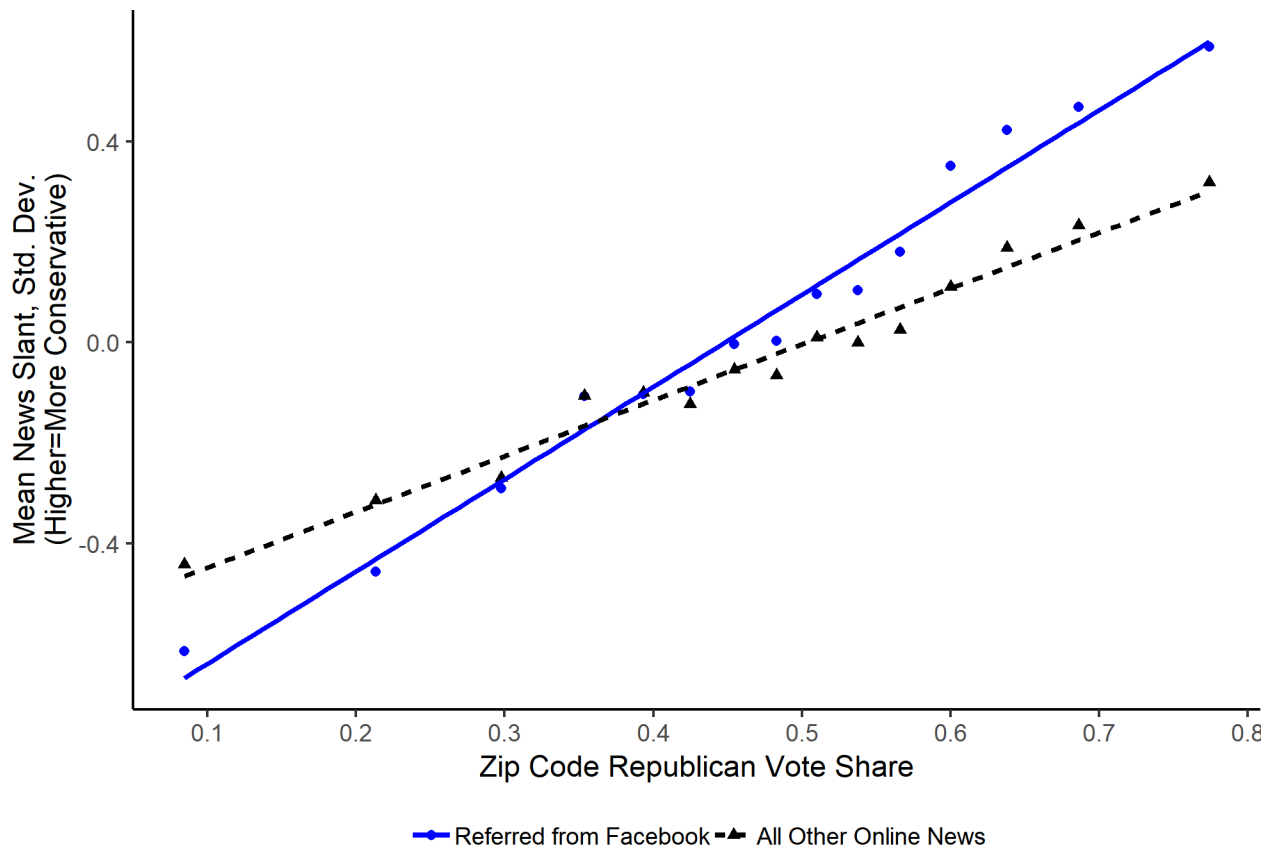
- Abramowitz, A. I. and S. Webster (2016). The Rise of Negative Partisanship and the Nationalization of U. S. Elections in the 21st Century. *Electoral Studies* 41, 12–22.
- Allcott, H., L. Braghieri, S. Eichmeyer, and M. Gentzkow (2019). The Welfare Effects of Social Media. *NBER Working Paper*.
- Allcott, H. and M. Gentzkow (2017). Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives* 31(2), 211–236.
- Anderson, M. L. (2008). Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association* 103(484), 1481–1495.
- Anderson, S. P. and J. McLaren (2012). Media Mergers and Media Bias with Rational Consumers. *Journal of the European Economic Association* 10(4), 831–859.
- Angrist, J. D. and I. Fernandez-Val (2013). ExtrapoLATE-ing: External Validity and Overidentification in the LATE Framework. In D. Acemoglu, M. Arellano, and E. Dekel (Eds.), *Advances in Economics and Econometrics - Tenth World Congress*, pp. 401–433.
- Ansolabehere, S. and J. Rodden (2012). Harvard Election Data Archive.
- Aronow, P. M. and A. Carnegie (2013). Beyond LATE: Estimation of the Average Treatment Effect with an Instrumental Variable. *Political Analysis* 21(04), 492–506.
- Bail, C., L. Argyle, T. Brown, J. Bumpus, H. Chen, M. Hunzaker, J. Lee, M. Mann, F. Merhout, and A. Volfovsky (2018). Exposure to Opposing Views can Increase Political Polarization: Evidence from a Large-Scale Field Experiment on Social Media. *Proceedings of the National Academy of Sciences of the United States of America*.
- Bakshy, E., S. Messing, and L. A. Adamic (2015). Exposure to Ideologically Diverse News and Opinion on Facebook. *Science* 348(6239), 1130–1132.
- Barberá, P. (2015). How Social Media Reduces Mass Political Polarization. Evidence from Germany, Spain, and the U.S. *Job Market Paper*, 44.
- Berinsky, A. J., G. A. Huber, and G. S. Lenz (2012). Evaluating Online Labor Markets for Experimental Research: Amazon.com’s Mechanical Turk. *Political Analysis* 20(3), 351–368.
- Boxell, L., M. Gentzkow, and J. M. Shapiro (2017). Is the Internet Causing Political Polarization? Evidence from Demographics. *NBER Working Paper*.
- Chen, Y. and D. Y. Yang (2018). The Impact of Media Censorship: Evidence from a Field Experiment in China. *Job Market Paper*.

- Comscore (2017). Comscore Web Behavior Database.
- Coppock, A., E. Ekins, and D. Kirby (2018). The Long-lasting Effects of Newspaper Op-Eds on Public Opinion. *Quarterly Journal of Political Science* 13(1), 59–87.
- DellaVigna, S. and E. Kaplan (2007). The Fox News Effect: Media Bias and Voting. *The Quarterly Journal of Economics* 122(3), 1187–1234.
- Druckman, J. N. and M. S. Levendusky (2019). What do we Measure when we Measure Affective Polarization? *Public Opinion Quarterly* 83(1), 114–122.
- Flaxman, S. R., G. Sharad, and J. M. Rao (2016). Filter Bubbles, Echo Chambers, and Online News Consumption. *Public Opinion Quarterly* 80, 298–320.
- Gentzkow, M., J. M. Shapiro, and D. F. Stone (2015). Media Bias in the Marketplace: Theory. In *Handbook of Media Economics, 1B*, Volume 1, pp. 623–645. Elsevier B.V.
- Gentzkow, M. (2006). Television and Voter Turnout. *The Quarterly Journal of Economics* 121(3), 931–972.
- Gentzkow, M. (2016). Polarization in 2016. *Toulouse Network for Information Technology Whitepaper*, 1–22.
- Gentzkow, M. and J. M. Shapiro (2006). Media Bias and Reputation. *Journal of Political Economy* 114(2), 280–316.
- Gentzkow, M. and J. M. Shapiro (2010). What Drives Media Slant? Evidence From U.S. Daily Newspapers. *Econometrica* 78(1), 35–71.
- Gentzkow, M. and J. M. Shapiro (2011). Ideological Segregation Online and Offline. *Quarterly Journal of Economics* 126(4), 1799–1839.
- Gerber, A. S., D. Karlan, and D. Bergan (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.
- Gift, K. and T. Gift (2015). Does Politics Influence Hiring? Evidence from a Randomized Experiment. *Political Behavior* 37(3), 653–675.
- Gorodnichenko, Y., T. Pham, and O. Talavera (2018). Social Media, Sentiment and Public Opinions: Evidence from #BREXIT and #USELECTION. *NBER Working Paper*.
- Gosling, S. D., P. J. Rentfrow, and W. B. Swann (2003). A Very Brief Measure of the Big-Five Personality Domains. *Journal of Research in Personality* 37(6), 504–528.
- Guess, A., B. Nyhan, B. Lyons, and J. Reifler (2018). Avoiding the Echo Chamber about Echo Chambers.

- Guess, A., B. Nyhan, and J. Reifler (2017). "You're Fake News" Findings from the Poynter Media Trust Survey.
- Guess, A. M. (2018). (Almost) Everything in Moderation: New Evidence on Americans' Online Media Diets. *Working Paper*.
- Halberstam, Y. and B. Knight (2016). Homophily, group size, and the diffusion of political information in social networks: Evidence from Twitter. *Journal of Public Economics* 143, 73–88.
- Heckman, J. J., S. Urzua, and E. J. Vytalil (2006). Understanding Instrumental Variables in Models With Essential Heterogeneity. *The Review of Economics and Statistics* 88(August), 389–432.
- Hortacsu, A., M. R. Wildenbeest, and B. De Los Santos (2012). Testing Models of Consumer Search using Data on Web Browsing and Purchasing Behavior. *American Economic Review* 102, 2955–2980.
- Iyengar, S. and K. S. Hahn (2009). Red Media, Blue Media: Evidence of Ideological Selectivity in Media Use. *Journal of Communication* 59(1), 19–39.
- Iyengar, S. and M. Krupenkin (2018). The Strengthening of Partisan Affect. *Political Psychology* 39, 201–218.
- Iyengar, S., Y. Lelkes, M. Levendusky, N. Malhotra, and S. J. Westwood (2019). The Origins and Consequences of Affective Polarization in the United States. *Annual Review of Political Science* 22(1), 129–146.
- Jo, D. (2018). Better the Devil You Know: An Online Field Experiment on News Consumption. *Job Market Paper*.
- Lelkes, Y. (2016). The Polls-Review: Mass Polarization: Manifestations and Measurements. *Public Opinion Quarterly* 80, 392–410.
- Lelkes, Y., G. Sood, and S. Iyengar (2015). The Hostile Audience: The Effect of Access to Broadband Internet on Partisan Affect. *American Journal of Political Science* 61(1), 5–20.
- Martin, G. J. and A. Yurukoglu (2017). Bias in Cable News: Persuasion and Polarization. *American Economic Review*.
- Mason, L. (2015). "I Disrespectfully Agree": The Differential Effects of Partisan Sorting on Social and Issue Polarization. *American Journal of Political Science* 59(1), 128–145.
- McConnell, C., Y. Margalit, N. Malhotra, and M. Levendusky (2018). The Economic Consequences of Partisanship in a Polarized Era. *American Journal of Political Science* 62(1), 5–18.
- Mummolo, J. and C. Nall (2016). Why Partisans Do Not Sort: The Constraints on Political Segregation. *The Journal of Politics* 79(1), 45–59.

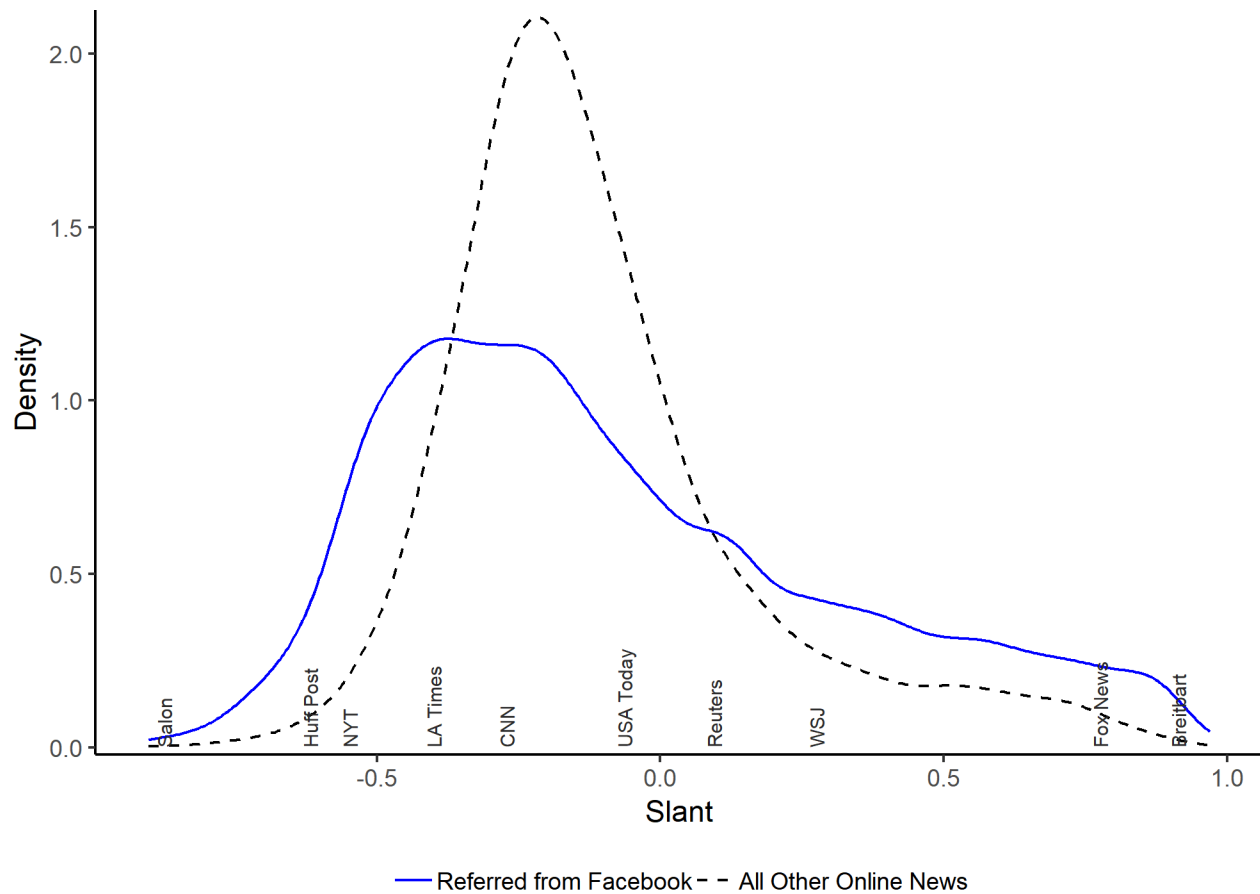
- Peterson, E., G. Shared, and S. Iyengar (2018). Echo Chambers and Partisan Polarization: Evidence from the 2016 Presidential Campaign. *Working Paper*.
- Pew (2014). *Political Polarization and Media Habits*. Pew Research Center.
- PEW (2019). Public Highly Critical of State of Political Discourse in the U.S. Technical report.
- Rand, D. G., A. Peysakhovich, G. T. Kraft-Todd, G. E. Newman, O. Wurzbacher, M. A. Nowak, and J. D. Greene (2014). Social Heuristics Shape Intuitive Cooperation. *Nature Communications* 5, 1–12.
- Reit, E., R. Willer, and J. Zaki (2017). Causes and Consequences of Political Empathy. *Work in Progress, Stanford University*.
- Reuters Institute (2019). Digital News Report 2019.
- Schroeder, E. and D. F. Stone (2015). Fox News and political knowledge. *Journal of Public Economics* 126, 52–63.
- Shane, F. (2005). Cognitive Reflection and Decision Making Author(s):. *The Journal of Economic Perspectives* 19(4), 25–42.
- Shearer, E. and K. E. Matsa (2018). News Use Across Social Media Platform 2018. *PEW Research Center*.
- Strömberg, D. (2015). Media and Politics. *Annual Review of Economics* 7(1), 173–205.
- Sunstein, C. (2017). *#Republic: Divided Democracy in the Age of Social Media*. Princeton University Press.
- Tucker, J. A., A. Guess, P. Barbera, C. Vaccari, A. Siegel, S. Sanovich, D. Stukal, and B. Nyhan (2018). Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature.
- Tufekci, Z. (2015). Algorithmic Harms beyond Facebook and Google: Emergent Challenges of Computational Agency. *Journal on Telecommunications & High Tech Law* 13(23), 203–216.
- Yeager, D. S., J. A. Krosnick, L. Chang, H. S. Javitz, M. S. Levendusky, A. Simpser, and R. Wang (2011). Comparing the Accuracy of RDD Telephone Surveys and Internet Surveys Conducted with Probability and Non-Probability Samples. *Public Opinion Quarterly* 75(4), 709–747.
- Zuiderveen Borgesius, F., D. Trilling, J. Möller, B. Bodó, C. De Vreese, and N. Helberger (2016). Should We Worry about Filter Bubbles? *Internet Policy Review* 5(1), 1–16.

Figure 1: Ideology and Slant of News Consumption



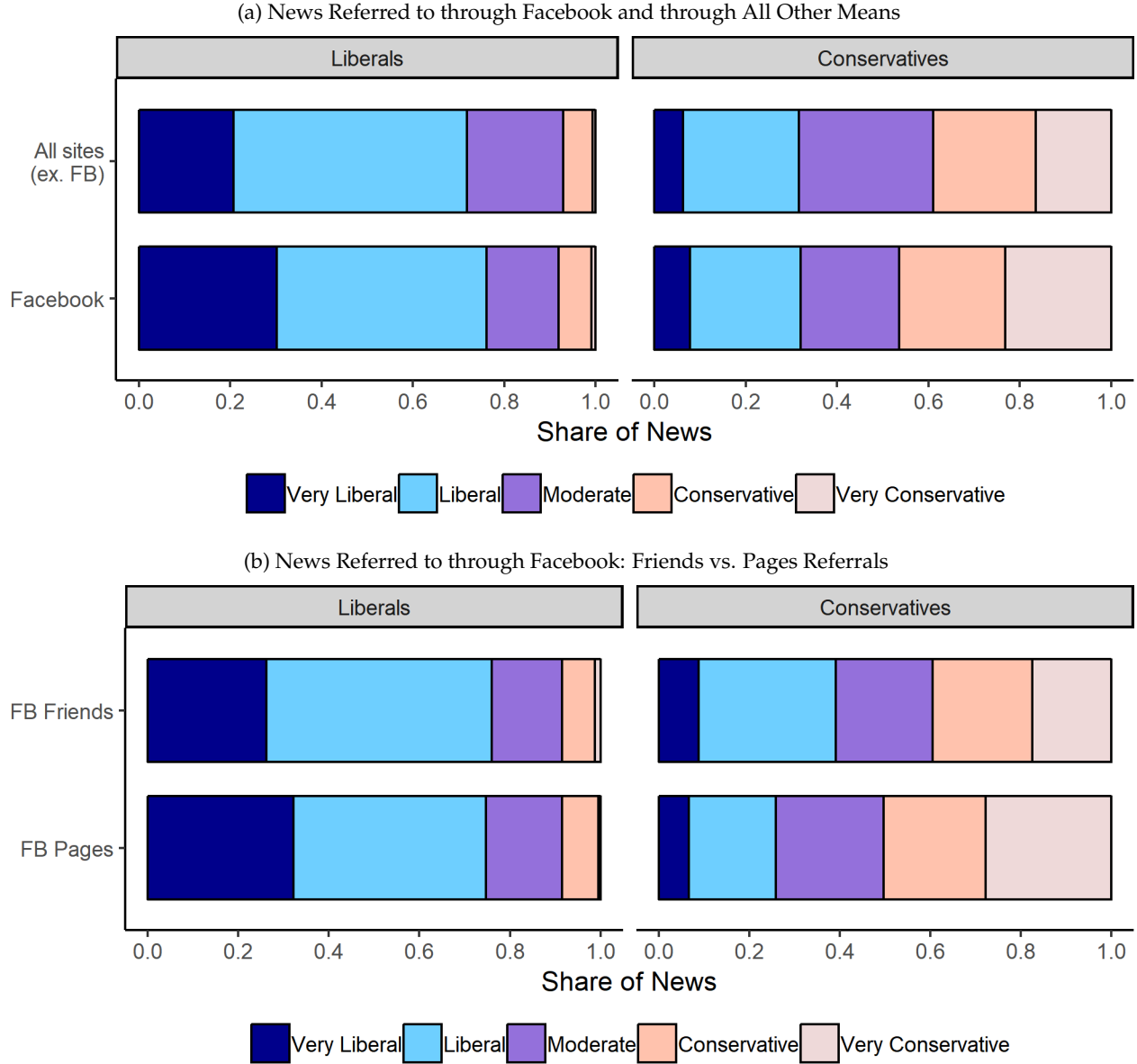
This figure shows the correlation between political ideology and online news consumption. It presents a binned scatter plot based on 2017 Comscore data. The Republican vote share in the x-axis is based on 2008 zip code level voting data (Mummolo and Nall, 2016). The mean slant in the y-axis is calculated as the mean slant of all news sites visited, where the slant of each domain is based on Bakshy et al. (2015). A visit to a news site is referred from Facebook if the referring domain is "facebook.com". The sample includes all users who visited at least two news sites through Facebook and two news sites through other means.

Figure 2: Distribution of Mean News Slant



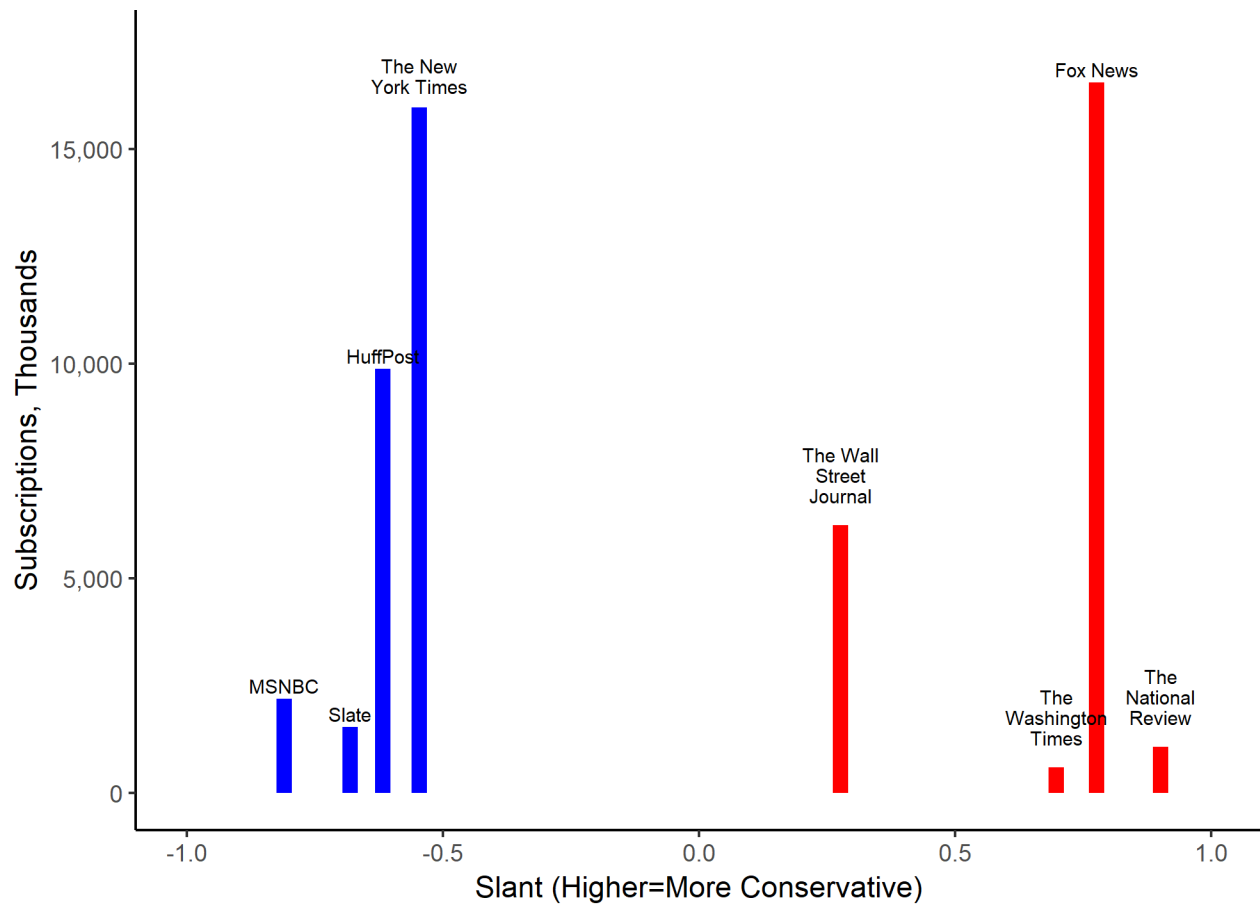
This figure shows the distribution of the mean slant of news consumed by individuals based on 2017 Comscore data. The slant of each domain is based on Bakshy et al. (2015). A visit to a news site is referred from Facebook if the referring domain is “facebook.com”. The sample includes all users who visited at least two news sites through Facebook and two news sites through other means. Major news outlets are added to the x-axis for reference.

Figure 3: Referral to News Sites, Control Group Browser Extension Data



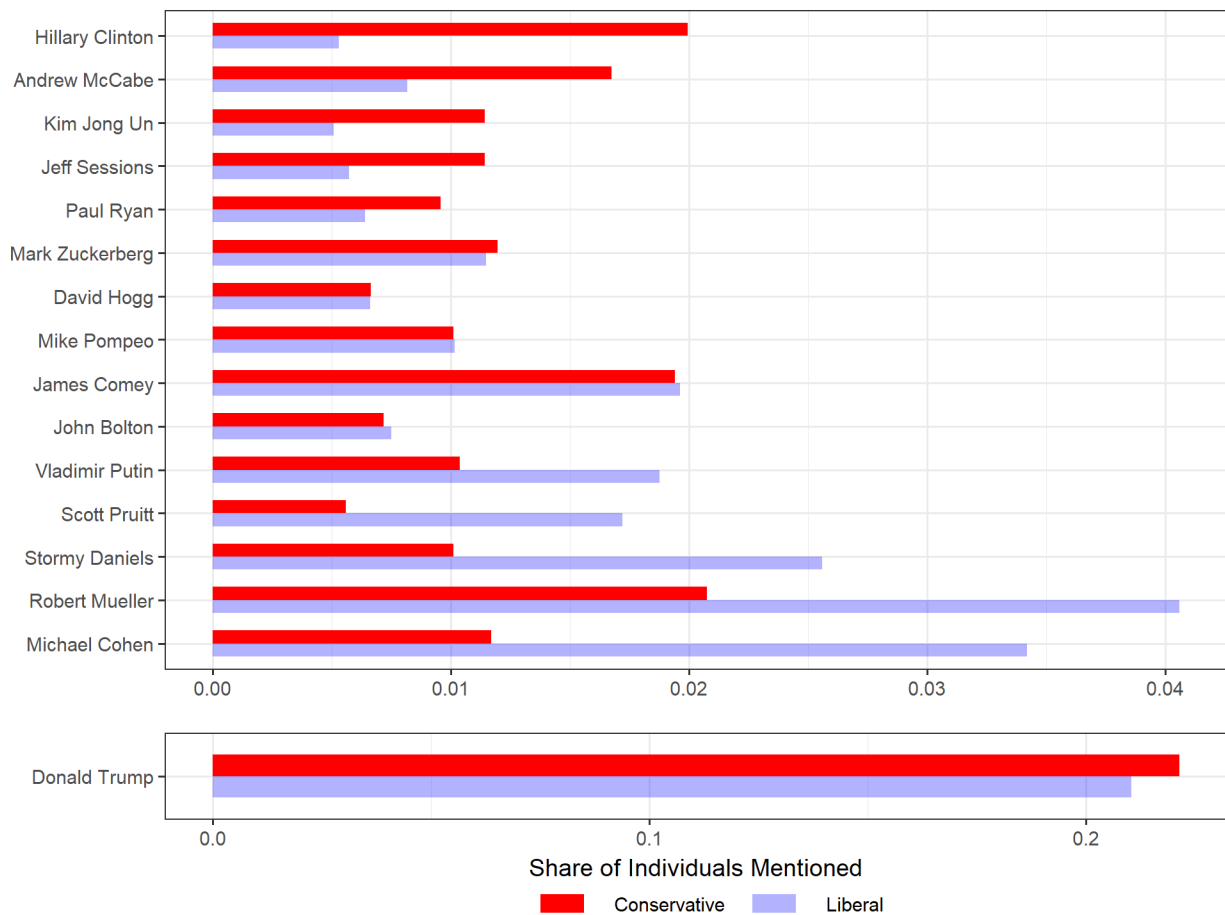
These figures show the mean distribution of news consumed online from leading outlets according to the consumer's ideology. The share of news consumed for each category is first calculated at the individual level and then the mean is calculated for each group. For more details on processing leading outlets see Appendix A.3. The first sub-figure shows all news consumed when a user clicked any link in Facebook compared to news consumed through any other means. The second sub-figure focuses on news consumed through Facebook and compared news sites accessed by clicking links shared by Facebook friends and news sites accessed by clicking links shared by Facebook pages. The figure is based on the browser extension data for participants in the control group who kept the extension installed for at least two weeks. Only participants who visited at least one news site by clicking a link shared by a Facebook page, one news site by clicking a link shared by a friend on Facebook and one news site through other means are included. To increase power, the figure is based on one month of extension data.

Figure 4: Primary Assigned Outlets



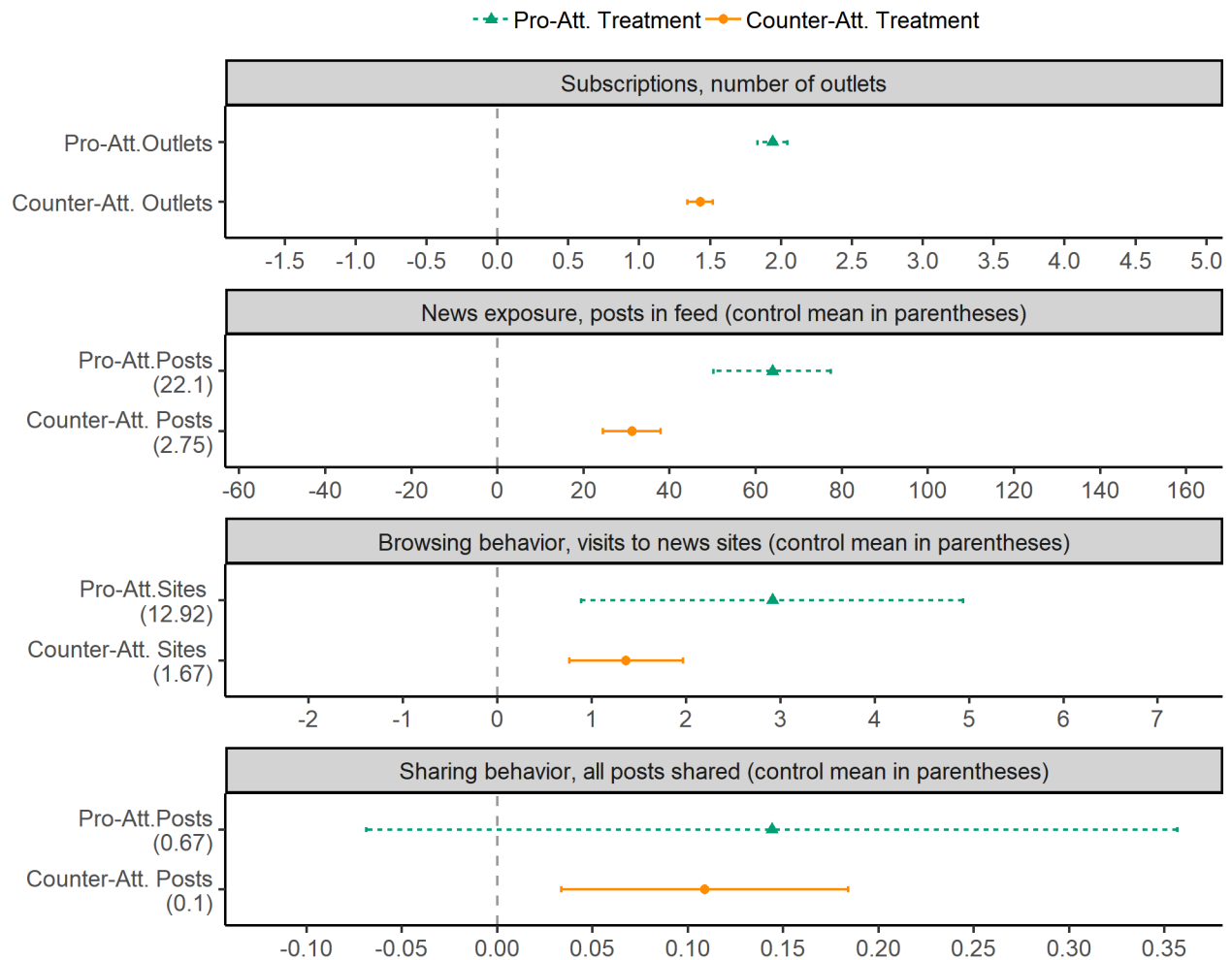
This figure displays the primary liberal and conservative outlets offered in the experiment. The x-axis is the slant of the outlets, as determined by Bakshy et al. (2015), and the y-axis is the total number of individuals who have subscribed to each outlet on Facebook in April 2018.

Figure 5: Figures Discussed in the News During the Study Period



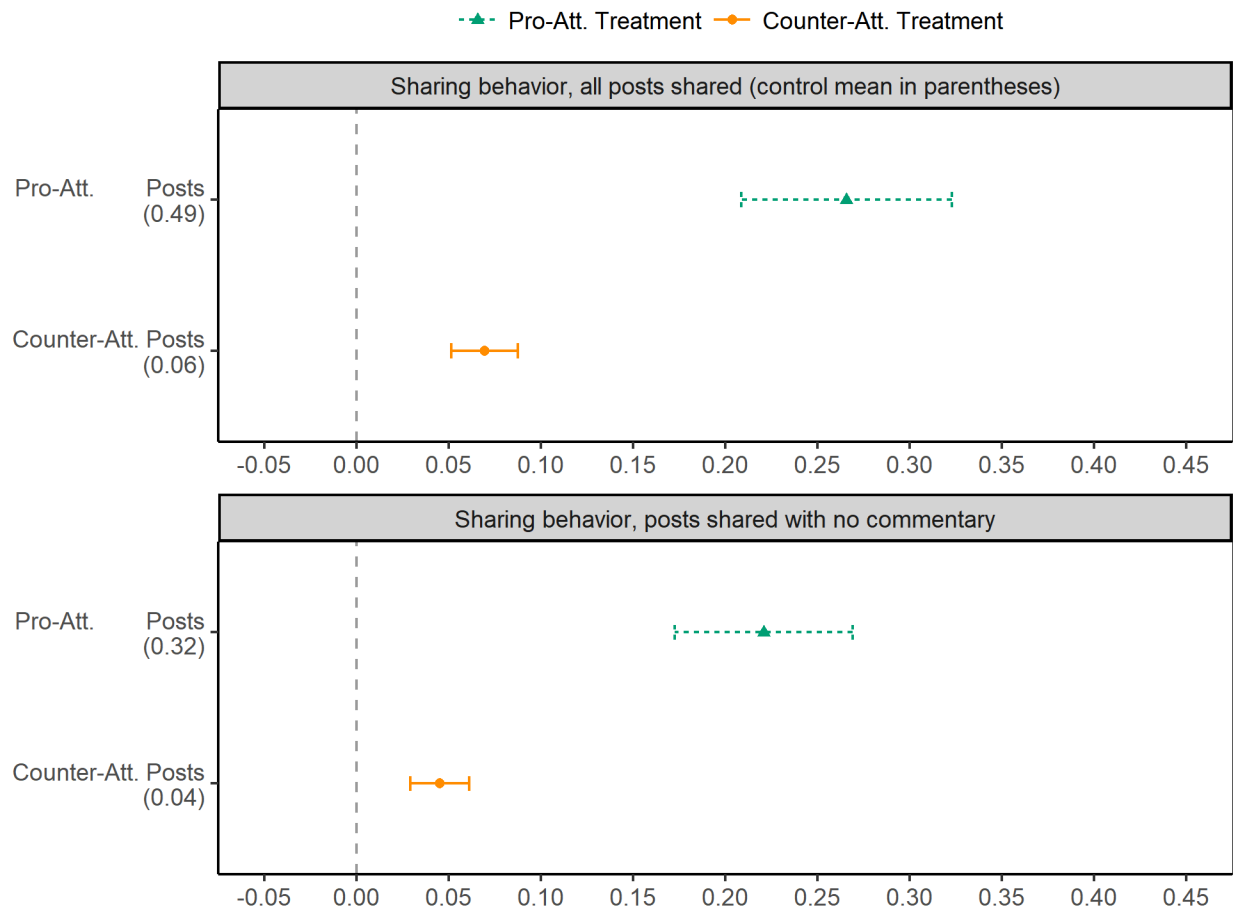
This figure shows the prominent men and women mentioned in posts shared by the primary outlets between February 28 and April 25, the median dates the baseline survey and endline survey were taken. The x-axis is the share of times an individual was mentioned in a post by one of the four primary conservative outlets (top bars) and by one of the four primary liberal outlets (bottom bars), of all individuals mentioned. To fit all the figures on the same scale, the x-axis is broken for Donald Trump who is by far the most dominant figure mentioned. The figures were identified using the Spacy Natural Language Processing algorithm. To simplify the graph, the names 'Trump' and 'Donald Trump' were determined to be the same individual, even though 'Trump' could refer to other members in President Trump's family.

Figure 6: Effects of the Pro- and Counter-attitudinal on Subscriptions, News Exposure, News Sites Visited and Sharing Behavior, Two Weeks Following the Intervention



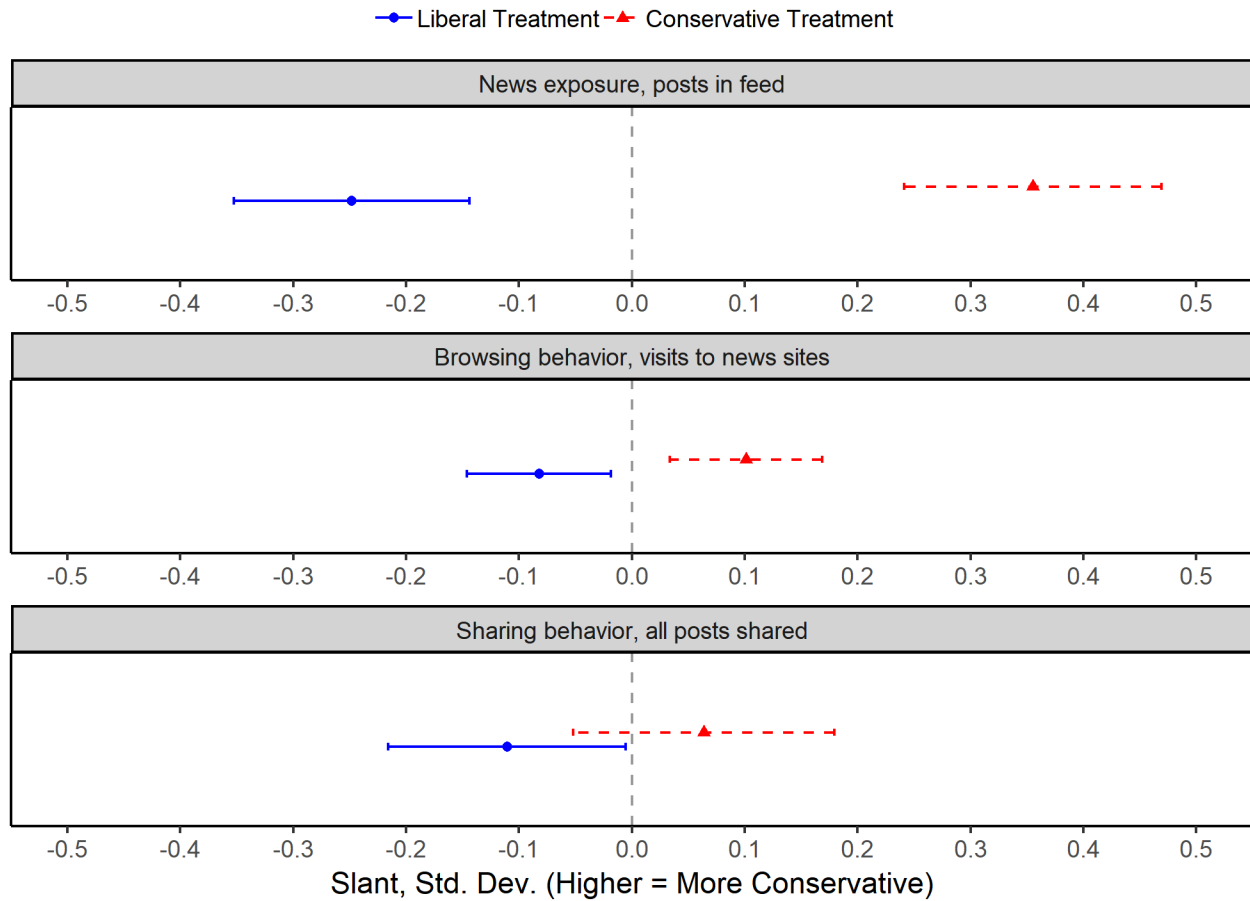
This figure shows the effect of the pro-attitudinal and counter-attitudinal treatments on engagement with each individual's potential outlets. Each row in the figure is estimated by regressing engagement with the four potential pro-attitudinal outlets or four potential counter-attitudinal outlets on the treatment. The outcomes are the number of outlets individuals subscribed to, the number of posts from the outlets that appeared in their feed, the number of times they visited the outlets' websites and the number of posts shared from the outlets. Data based on 1,703 participants who installed the extension and provided permissions to access their posts for at least two weeks following the intervention. The regressions control for the outcome measure in baseline if it exists. Error bars reflect 90 percent confidence intervals.

Figure 7: Effects of the Pro- and Counter-attitudinal Treatments on Number of Posts Shared, Access Posts Sub-Sample



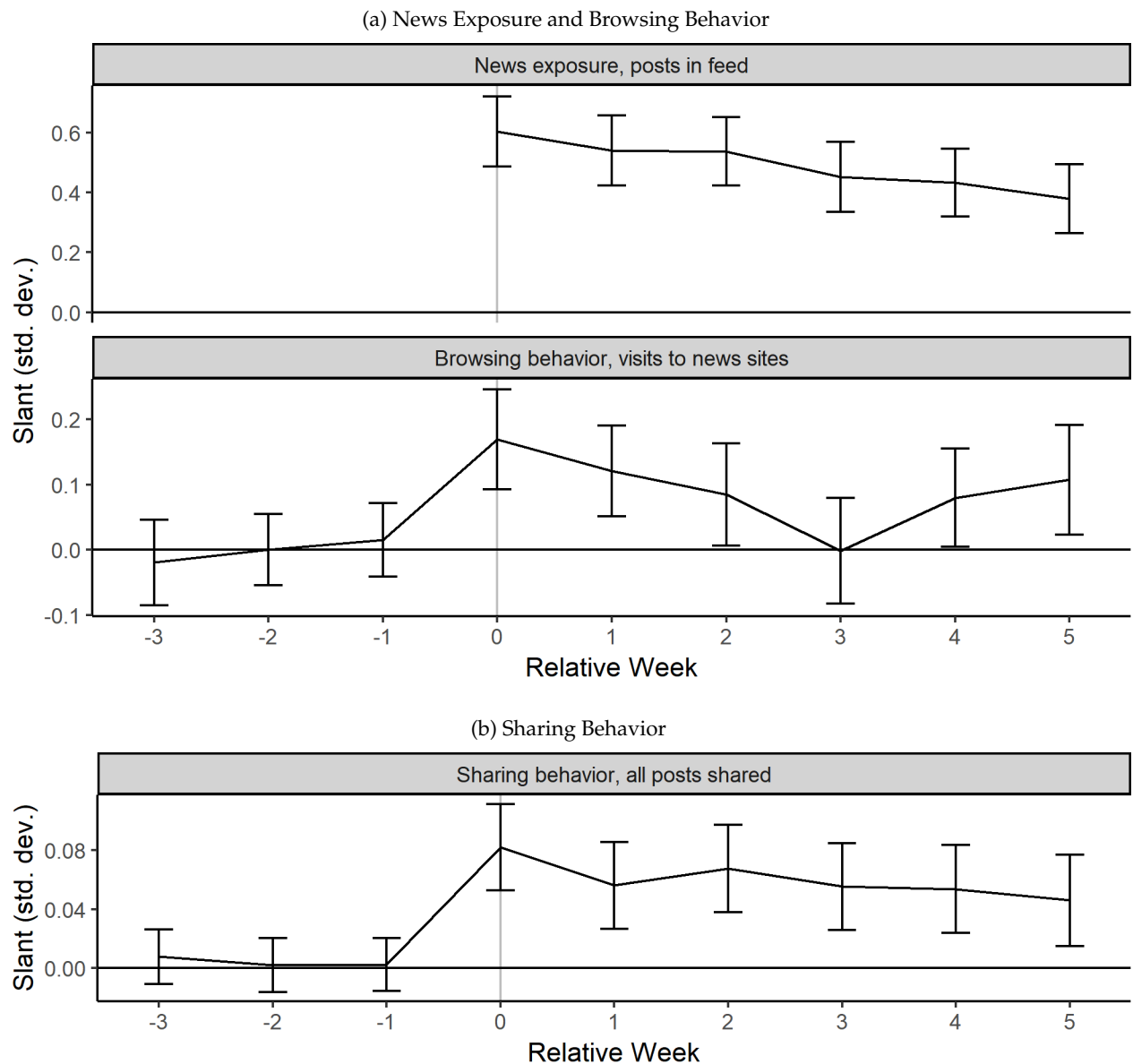
This figure shows the effect of the pro-attitudinal and counter-attitudinal treatments on engagement with each individual's four potential pro-attitudinal outlets and four potential counter-attitudinal outlets. Each row in the figure is estimated by regressing engagement with the four potential pro-attitudinal outlets or four potential counter-attitudinal outlets on the treatment. The outcomes in both panels are the number of posts the individuals shared from these outlets. The first panel includes all posts and the second panel includes only posts that were shared without any commentary by the participant. Data based on the access posts sub-sample: 34,575 participants who provided access to their posts for at least two weeks following the intervention. The regressions control each outcome measure in baseline. Error bars reflect 90 percent confidence intervals.

Figure 8: Effect of the Treatment on Slant of News Consumption



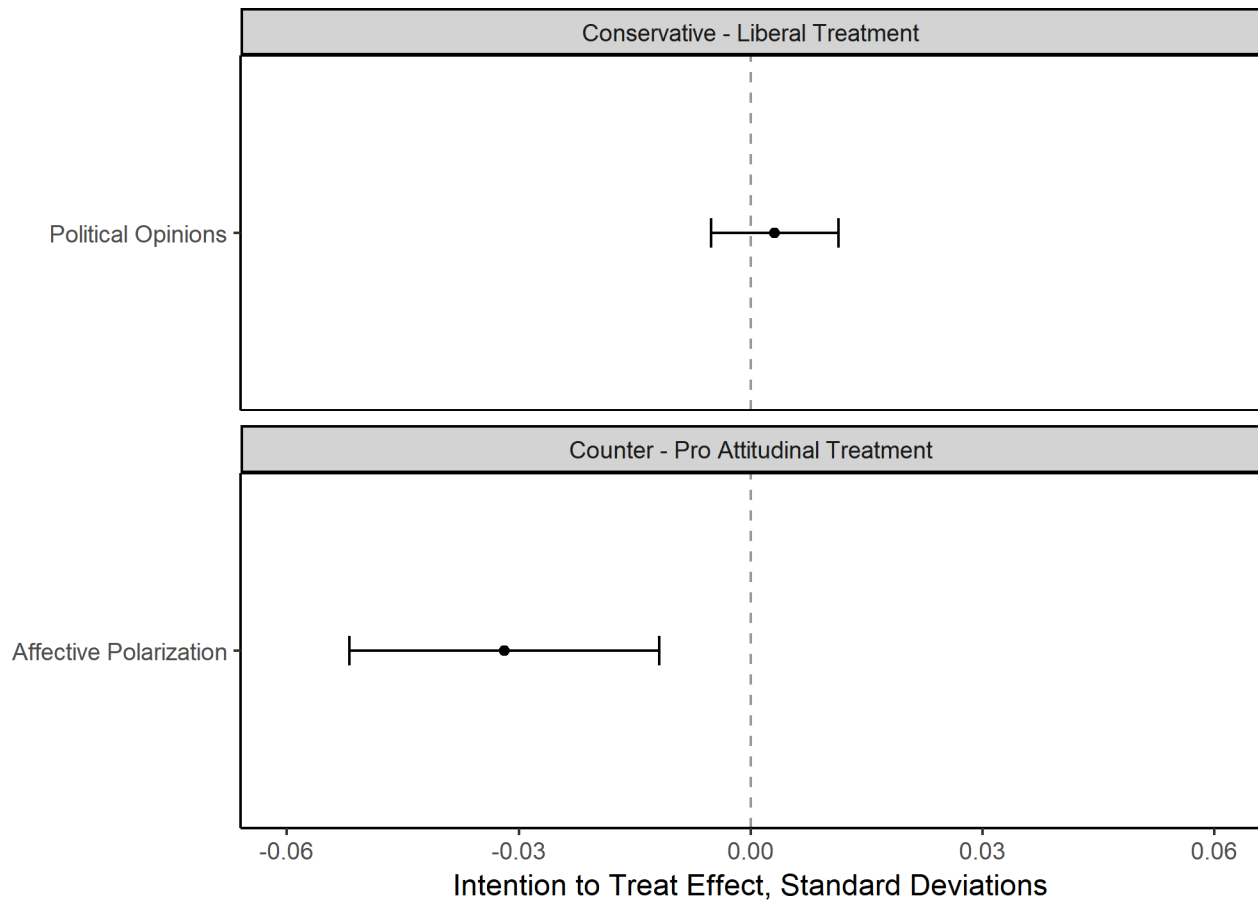
This figure shows the effect of the liberal and conservative treatments on the mean slant, in standard deviations, of all news individuals engaged with. Each panel in the figure is estimated by regressing the slant of outlets on the treatment. The regressions control for the outcome in baseline if it exists. The figure displays the slant for three outcomes: Exposure to posts on Facebook (panel 1), news sites visited (panel 2) and posts shared (panel 3). Data based on 1,703 participants who installed the extension and provided permissions to access their posts for at least two weeks following the intervention. Error bars reflect 90 percent confidence intervals.

Figure 9: Mean Slant over Time, Conservative Treatment, Compared to Liberal Treatment



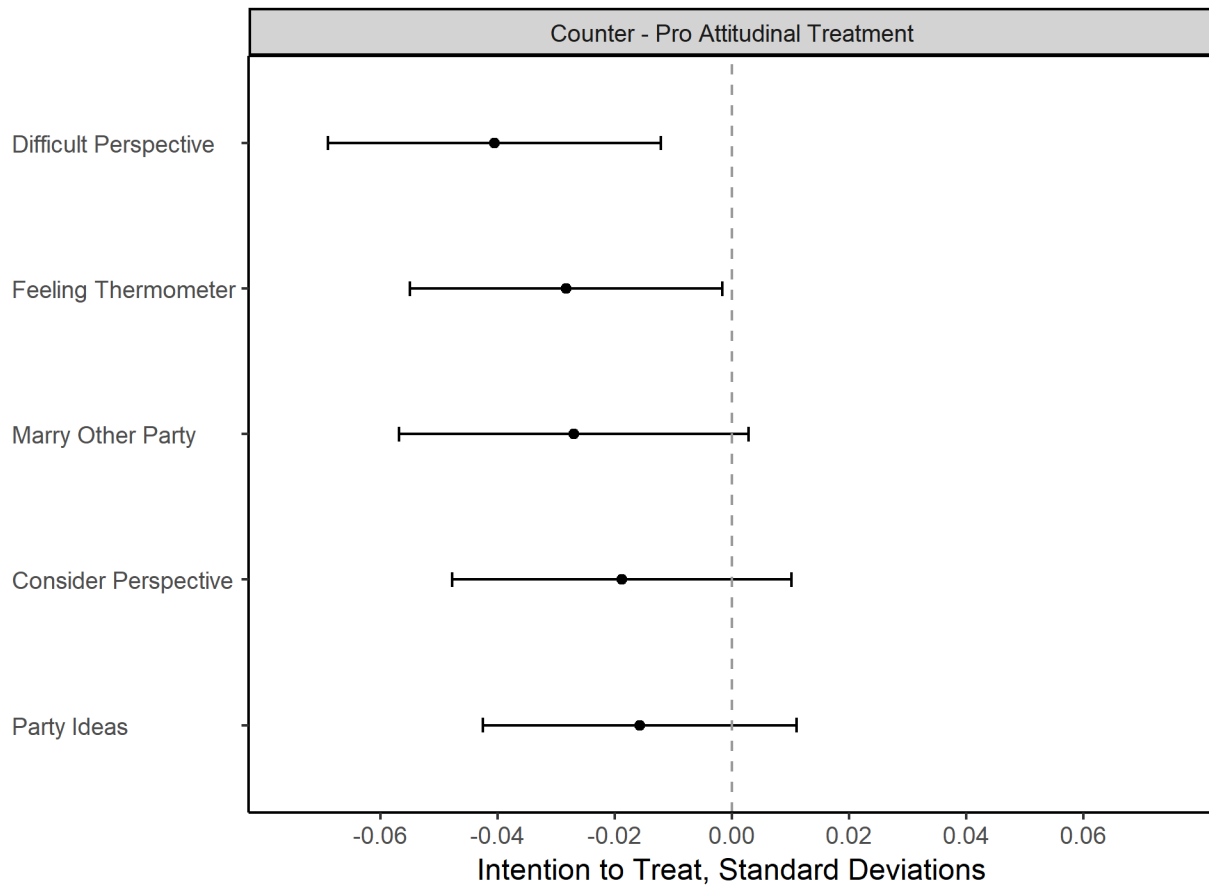
These figures show the difference between the effect of the liberal and conservative treatments on the mean slant over time. Each panel presents a series of regressions, estimated by regressing the slant of outlets on the treatment in a specific week. The regression control for the outcome in baseline when it exists. In the first figure the data is based on 1,597 participants who kept the extension installed for at least six weeks following the intervention. In the second figure the data is based on 29,115 participants who provided access to posts they shared for at least six weeks following the intervention. Error bars reflect 90 percent confidence intervals.

Figure 10: Effect of the Treatment on Political Opinions and Polarization



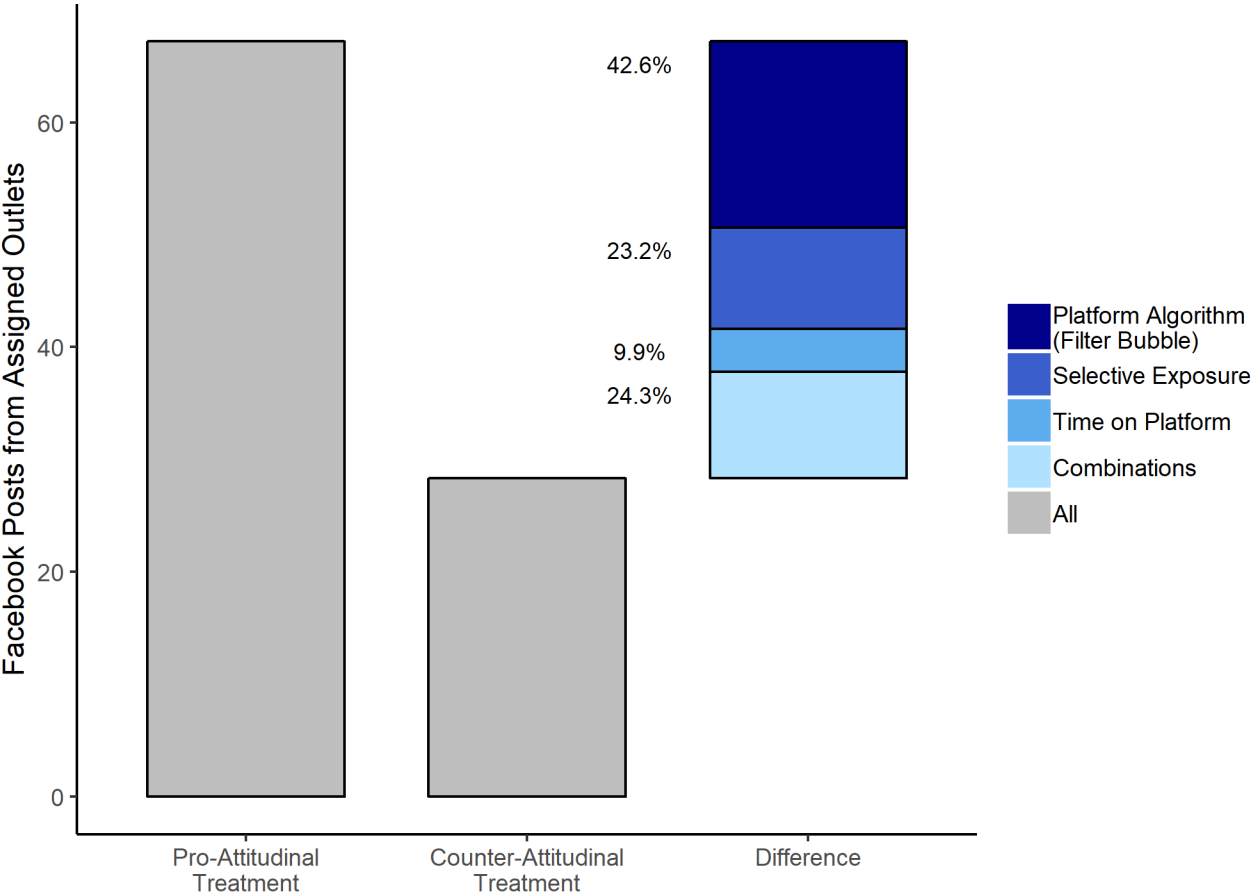
The first panel shows the effect of the conservative treatment on the political opinions index, compared to the liberal treatment. A higher value is associated with a more conservative outcome. The second panel shows the effect of the pro-attitudinal treatment on the affective polarization index, compared to the counter-attitudinal treatment. A higher value is associated with a more polarized outcome. The indices are described in section 3.4.2 and the regressions specifications are detailed in section 3.6. Error bars reflect 90 percent confidence intervals.

Figure 11: Effect of the Treatment on Affective Polarization - Individual Measures



Each row represents the result of a regression estimating the effect of the treatment on the dependent variables composing the affective polarization index. *Difficult Perspective* and *Consider Perspective* measures political empathy (Reit et al., 2017). The former is the difference in how difficult it is to see things from each party's point of view, and the latter variable is the difference in how important it is to consider the perspective of each party. *Feeling Thermometer* is the difference in a feeling thermometer question asking participants how warm they feel toward each party. *Marry Opposing Party* is how participants would feel if their son/daughter married someone from the opposing party. *Party Ideas* is the difference in how many good ideas each party has. The outcomes are described in more detail in section 3.4.2. The regressions specifications are detailed in section 3.6. Error bars reflect 90 percent confidence intervals.

Figure 12: Decomposing the Gap Between the Effect of the Pro-attitudinal and Counter-attitudinal Treatments on Exposure



This figure decomposes the gap between the number of posts from potential pro-attitudinal outlets participants in the pro-attitudinal treatment were exposed to and the number of posts from potential counter-attitudinal outlets participants in the counter-attitudinal treatment were exposed to. The y-axis is the number of posts seen per day and the x-axis is the treatment. The terms are described in Section 6 and the calculations appear in Appendix Table A.15.

Table 1: The Association Between the Interaction of News Consumption Source and Consumer's, and the Slant of News Consumed

	Slant	Slant	Mean Slant	Mean Slant	Mean Slant
	(1)	(2)	(3)	(4)	(5)
Rep. Vote Share	0.821*** (0.097)		0.213*** (0.027)		
FB References * Rep. Vote Share	0.932*** (0.182)	0.415*** (0.075)			
FB Share of News Ref. * Rep. Vote Share			2.603*** (0.211)	0.874*** (0.106)	
FB Share of Visits * Rep. Vote Share					1.213*** (0.451)
Unit of Observation	Site	Site	Ind.	Ind.*Month	Ind.*Month
Individual FE		X		X	X
Month FE		X		X	X
Demographics			X		
Observations	2,181,674	2,181,674	57,839	263,285	263,285

This table shows the association between ideology, social media and the slant of news consumed. The 2008 Republican vote share is determined by each individual's zip code. The slant of outlets is based on Bakshy et al. (2015). In columns (1)-(2) each observation is a website visited, *FB Reference* refers to a visit where the referring domain is "facebook.com" and the dependent variable is the mean slant of sites visited calculated based on Bakshy et al. (2015), where a higher slant is more conservative. In column (3) each observation is an individual and in columns (4)-(5) each observation is an individual*month. The dependent variable is measured in standard deviations. In columns (3)-(5) the dependent variable is the mean slant of all sites visited. *FB Share of News Ref* refers to the share of news sites an individual visited through Facebook and *FB Share of Visits* refers to the share of visits to Facebook among all websites visited. The sample includes all users who visited at least two news sites through Facebook and two news sites through other means.

Table 2: Samples, Data Sources and Outcomes

Sample / sub-sample	Data sources	Number of Participants	Main Outcomes Measured
Baseline sample	Baseline survey. Facebook data on participants' subscriptions to outlets.	37,492 (all participants)	Compliance - subscription to outlets in the intervention.
Extension sub-sample	Browser data from participants who installed the chrome extension for at least two weeks.	1,839	Exposure - posts seen on the Facebook feed. Browsing behavior - news sites visited.
Access posts sub-sample	Facebook data on posts shared by participants who provided permissions to access their posts for at least two weeks	34,575	Sharing behavior. Subscription to outlets over time.
Endline survey sub-sample	Baseline and endline surveys of participants who completed both surveys.	17,634	Political opinions. Affective polarization.

This table describes the main sample and sub-samples used in the analysis along with the data source, the number of participants and the main outcomes analyzed. The sub-samples and data are described in section 3.3. The outcomes are described in section 3.4.

Table 3: Balance Table by Assignment to the Liberal and Conservative Treatments

Variable	Mean		Difference Between Treatments		
	Sample	US	Control - Lib.	Control - Cons.	Cons. - Lib.
Baseline Survey					
Ideology (-3, 3)	-0.61	0.15	0.013	0.012	0.001
Republican	0.17	0.25	-0.006	0.001	-0.007
Independent	0.37	0.29	-0.004	-0.003	-0.000
Democrat	0.38	0.3	0.009	0.003	0.006
Vote Support Clinton	0.53		-0.004	-0.003	-0.001
Vote Support Trump	0.26		0.004	-0.001	0.005
Feeling Therm, Rep.	29.1	43.1	0.126	0.253	-0.127
Feeling Therm, Dem.	47.0	48.7	0.375	0.449	-0.075
Difficult Pers., Rep. (1, 5)	3.13		0.018	0.003	0.015
Difficult Pers., Dem. (1, 5)	2.39		-0.000	0.006	-0.006
Most News Social	0.18	0.13	-0.002	0.002	-0.004
Device					
Took Survey Mobile	0.67		-0.011*	-0.000	-0.011*
Facebook					
Female	0.52	0.51	-0.005	-0.002	-0.004
Age	48.5	47.5	0.229	-0.216	0.446*
Total Subscriptions	474		5.419	9.207	-3.788
News Outlets Subscriptions	8.47		-0.081	-0.075	-0.006
News Outlets Slant (-1, 1)	-0.20		0.002	0.001	0.001
Access Posts, Pre-Treatment	0.98		0.001	0.005***	-0.004**
Attrition					
Took Followup Survey	0.47		0.028***	0.028***	-0.001
Access Posts, 2 Weeks	0.92		0.001	0.008**	-0.007**
Ext Install, 2 Weeks	0.05		0.002	-0.002	0.004
F-Test			1.266	0.825	1.082
P-value			[0.161]	[0.724]	[0.351]

This table presents descriptive statistics, along with the difference between participants assigned to the liberal treatment, conservative treatment, and control group. *Ideology* is mean self-reported ideology on a seven-point scale. *Republican*, *Independent*, and *Democrat* are party affiliation, including leaners. *Vote Support* is the share of participants who voted for the candidate or did not vote and supported the candidate. Other options include not supporting any candidate or voting for a third candidate. *Feeling Therm.* is the feeling thermometer score on a 0-100 degree scale. *Difficult Pers.* is whether participants find it difficult to see things from Democrats/Republicans point of view. *Most News* is where participants report getting most of their news. Other options include print, radio, web, TV and do not know. *Took Survey Mobile* is whether participants took the survey on a mobile phone. *Total Subscriptions* is the number of Facebook pages participants subscribed to in baseline. *News Outlets Subscriptions* is the number of subscriptions among leading news outlets. *News Outlets Slant* is the slant of news outlets subscriptions. *Access Posts* is whether participants provided access to the posts they shared. F-tests are calculated by regressing the treatment on all pre-treatment variables in the table, with missing values replaced with a constant and an indicator for a missing value. *p<0.1 **p<0.05 ***p<0.01

Table 4: Effect of the Social Media Feed on News Sites Visited

	Slant of News Sites (1)	Slant of News Sites Visited through FB (2)
Slant of FB Feed	0.314*** (0.069)	0.717*** (0.087)
Controls	X	X
F Stat	61.21	71.94
Observations	1,525	1,204

This table shows the effect of the Facebook feed on news sites visited. The dependent variable is the mean slant of all news sites visited and the independent variable is the mean slant of the participant's Facebook feed, instrumented by the treatment. In column (2) the dependent variable is only the slant of news sites visited through Facebook. Robust standard errors. *p<0.1 **p<0.05 ***p<0.01

Table 5: Effect of Exposure to Pro and Counter Attitudinal News on Affective Polarization

(a) Cross Sectional Correlation in Control Group		
	OLS Affective Polarization (1)	OLS (2)
FB Counter-Att. Share, Std. Dev.	-0.385*** (0.051)	
FB Congruence Scale, Std. Dev.		0.397*** (0.053)
Data	Control Group	Control Group
Observations	353	353

(b) Causal Effect Based on Experimental Variation		
	IV Affective Polarization	
	(1)	(2)
FB Counter-Att. Share, Std. Dev.	-0.137** (0.066)	
FB Congruence Scale, Std. Dev.		0.113** (0.057)
Controls	X	X
First Stage F	66.22	64.83
Share of Correlation in Control Group	0.35	0.28
Observations	1,071	1,071

These tables measure the associated between exposure to pro- and counter-attitudinal news on affective polarization. The tables use two summary-statistics. *Counter-Att. Share* is the share of counter-attitudinal news the participant was exposed to in Facebook, calculated as the share of news from counter-attitudinal outlets among all pro-attitudinal and counter-attitudinal outlets. The *Congruence Scale* is the mean slant of all news exposed to in Facebook, multiplied by (-1) for liberal participants. Sub-table (a) presents the results of regressions run only among control group participants, where the dependent variable is the affective polarization index and the independent variables are the two summary statistics (with no controls). Sub-table (b) shows the results of IV regressions, where the dependent variable is the affective polarization index, the independent variables are the two summary statistics and the instrument is the treatment. The regressions control for the covariates specified in section 3.6. The row titled *Share of Correlation in Control Group* divides the causal effect by the correlation in the control group. Robust standard errors.

*p<0.1 **p<0.05 ***p<0.01

Table 6: Effect of the Treatment on Attitudes toward Own vs Opposing Party

	Attitude Own Party	Attitude Opposing Party
	(1)	(2)
Counter-Att. Treatment	0.001 (0.014)	0.031** (0.014)
Pro-Att. Treatment	0.009 (0.013)	-0.003 (0.014)
Pro - Counter	0.008 (0.014)	-0.034** (0.014)
Observations	16,894	16,894

This table presents the effect of the pro-attitudinal treatment and counter-attitudinal treatment on attitudes toward the party the participant is associated with and the opposing party. Participants whose ideological leaning is defined as liberal (based on self-reported ideology, party affiliation, and candidate supported) are assumed to be associated with the Democratic Party and participants whose ideological leaning is defined as conservative are assumed to be associated with the Republican Party. The outcome for each party is an index composed of the following four questions: the feeling thermometer, how difficult it is to see things from each party's point of view, how important it is to consider the perspective of the party, and whether the party has good ideas. The *Marry Opposing Party* question is not included since consumers were only asked how they would feel if their son/daughter married someone from the opposing party. The controls are specified in section 3.6. Robust standard errors. *p<0.1 **p<0.05 ***p<0.01

Appendix

A Data Collection and Processing

A.1 Comscore Data

The Comscore Web Behavior Database Panel is a subset of Comscore’s opt-in Media Matrix Panel (Comscore, 2017), which is weighted to represent the US Internet population. Each observation in the dataset is a domain visited by a computer and includes details on the referring domain, the time visited, the number of pages visited and online purchases. While Comscore attempts to uniquely identify individuals, it is still possible that multiple individuals use the same machine. I assume that each id represents a unique individual.

I include in the sample only individuals who visited at least two news sites through Facebook and two news sites through other means. This ensures that for every individual in the sample, I can calculate the slant of news sites visited through Facebook and the slant of other news sites visited, and thus any difference in the slant does not stem from differences in the individuals who visited the news sites. I require at least two news sites for each category to slightly decrease measurement error. The results are similar when including individuals who visit at least one news sites through Facebook and one news site through other means.

A.2 Leading News Outlets

Throughout the paper, I analyze participants’ engagement with leading outlets. The list of outlets and their slant are based on a dataset constructed by Bakshy et al. (2015). The authors use Facebook’s internal data and classify links to hard and soft news based on the words in the link. Hard news articles and opinions are related to issues including national news, politics or world affairs and soft news includes issues such as sports and entertainment. The alignment of each website is determined according to the self-reported ideology of Facebook users who share hard news links from the website.

Before using the dataset I removed the following popular domains which are not related to news outlets: Amazon, The White House, Twitter, Vimeo, Wikipedia, and YouTube.

I then remove the webreference (“www.”) from all outlets’ websites, so all outlets only contain the domain used.⁶¹ After processing the data, the list of leading outlets contains 488 websites. Similarly to Bakshy et al. (2015), I define very liberal outlets as outlets in the bottom quintile of the distribution of news slant outlets, liberal outlets as outlets in the second quintile, moderate

⁶¹Websites which appear twice in the dataset, with and without the webreference, are merged into one entry. For example, washingtonexaminer.com and www.washingtonexaminer.com are merged, with the slant defined as the mean slant of the two entries.

outlets as outlets in the middle quintile, conservative outlets as outlets in the fourth quintile and very conservative outlets are defined as outlets in the top quintile of the news slant distribution.

To measure the slant of Facebook posts, I rely on either the domain appearing in the post or the Facebook page that shared the post. Therefore, I match Facebook pages with the list of leading outlets by searching for all Facebook pages with names similar to the outlet's domain and manually checking the pages. Overall, Facebook pages were found for 374 out of 488 websites included from the Bakshy et al. (2015) dataset.

A.3 Surveys

A.3.1 Baseline Survey

The baseline survey took place from early February to mid-March 2018. Participants were recruited to the survey using Facebook ads. The ads either emphasized that a survey is being conducted and participants will take part in a gift card raffle or that the survey may be of interest for people who follow politics.⁶² The ads targeted Facebook users living in the US who are over 18 years old, and who are likely to click the ad and begin the survey. A subset of the ads targeted conservatives or moderate individuals who are often under-represented in Internet samples (Allcott and Gentzkow, 2017; Yeager et al., 2011). Since the majority of participants took the survey on a mobile phone, an additional subset of ads focused on desktop users, to ensure that a large enough sample of participants will be offered an option to install the Chrome extension. While the survey was open and participants could share the link or ad with anyone, the vast majority of participants entered the survey as a result of the ad.⁶³

40,514 participants took the survey and reached the screen where the intervention occurs. Of those participants, 37,492 are included in the final sample. Participants are excluded from the final sample for the following reasons: missing information on outlets the participant subscribes to either because the participant did not provide permissions to access that data or since the data was not collected properly in real-time (2.38%); participant already subscribed to too many of the outlets such that it was not possible to offer the participant four new liberal and four new conservative outlets (3.64%); technical issues with the Qualtrics survey which prevented some data from being collected (0.28%); or taking the survey a second time (0.01%). Finally, to ensure quality responses, participants who were likely to respond carelessly were also excluded. These

⁶²I do not find evidence for heterogeneous effects on political opinions or affective polarization according to the type of ad used.

⁶³To test whether many participants clicked the ad since someone shared it with them, I provided participants with a slightly modified link to the baseline survey after they completed the survey, and asked them to use this link if they wish to share the survey. Only 0.57% of participants entered the survey using this link. Any individual exposed to the ad could also share the ad or the link that appears in the ad with other individuals, and in such cases it is difficult to distinguish participants who reached the survey through an ad and those who reached it after the ad was shared with them. However, approximately 95% of exposures to the recruitment ad during the recruitment period were directly due to a sponsored ad appearing in one's Facebook feed and not due to someone sharing the ad. Therefore, it is likely that the vast majority of users entered the survey since a sponsored ad appeared in their feed.

include participants who completed all survey section until the intervention exceptionally quickly (in under three minutes where the median time was eleven minutes) and participants who did not answer at least half of the closed-ended, non-required questions (0.03%). The criteria determining whether to exclude a participant are all based on survey data submitted before the intervention occurs.

A.3.2 Endline Survey

Participants were invited to the endline survey between mid-April and early June 2018. Participants were mostly recruited to the survey using emails and Facebook ads.⁶⁴ To match endline survey responses with baseline survey responses, participants who began the survey were asked to log in to the endline survey through Facebook or supply an email address.⁶⁵ Endline responses are matched with baseline responses based on the following criteria: email address the survey invitation was sent to, Facebook id, email address entered in the survey, combination of zip code, first and last name if the combination is unique, and combination of first and last name if the combination is unique. 98.73% of the individuals who took the followup survey were matched.

19,693 participants began the endline survey and were matched with valid baseline survey responses. Respondents are included in the sample even if they did not complete the endline survey. If the same individual took the endline survey more than once, uncompleted surveys are excluded. If multiple observations still exist, only the first response is included for the individual. Overall 0.41% of valid matched responses were excluded as duplicates. 0.02% of responses were also excluded for taking the survey carelessly if the survey was completed exceptionally quickly (spent less than 20 seconds per survey page, compared to a median time of 67 seconds).

A.4 Facebook Data on Subscriptions and Posts Shared

I collect data on outlets participants subscribed to and posts they shared. Only posts shared by participants with their social networks of the following types are included in the analysis: status update, video, link or note (photos, albums, events, and music were excluded since these posts usually are not news-related). I exclude posts not matched with a news outlet in the analysis. I match Facebook posts to leading outlets based on the Facebook page which shared the posts and based on links contained in the post.

Matching links is based on the following steps:

⁶⁴A small share of participants were recruited through an invitation in the Chrome extension or through a Facebook notification.

⁶⁵Originally I planned to ask all participants to login to the survey using their Facebook account and this would have enabled me to perfectly match respondents in the baseline and endline surveys. However, the Cambridge Analytica scandal was exposed before the baseline and endline survey and as a result people were generally less inclined to use Facebook apps. Therefore, I provided respondents with an option to login using their email address.

1. For each outlet included in the experiment, a list of relevant domains is created. The list is created by checking which domains were shared by the Facebook page associated with the outlet and including the most dominant domains and any other domain directly linked to the outlet. For example, in addition to associating “huffingtonpost.com” with the Huffington Post, I associate “huffingtonpost.ca”, “huffpost.com” and other similar domains. For outlets not offered in the experiment, I define the main domain in the Bakshy et al. (2015) dataset as the only relevant domain.
2. If a link refers to a short alias, created by URL-shortening services such as tinyurl.com, it cannot be directly matched to an outlet based on the domain. Therefore, each URL is first converted to the final redirected URL before being matched to the list of domains.

A.5 Extension Data on Facebook Feed and Browsing Behavior

Data on news-related posts appearing in participants’ Facebook feeds is matched to outlets using the same method explained in the previous section. Similarly, news sites visited are matched to outlets based on the list of domains defined in the previous section. I exclude URLs that were visited for less than one second before another URL in the same domain was visited, as it is likely that the user did not actually observe the content of the website. If a URL is visited more than once within a 20-minute window, only the first visit is included.

A news site is determined to have been visited through Facebook if the website visited appeared in the participant’s Facebook feed in the 20 minutes proceeding to the website being visited.⁶⁶

B Additional Details on Empirical Strategy

B.1 Controls

To increase power, when estimating the effect on political opinion and affective polarization, I control for a set of pre-registered covariates. Appendix Table A.10 shows that the results regarding both political opinions and affective polarization are robust to excluding all covariates. I control for self-reported ideology, party affiliation, approval of President Trump, ideological leaning, age, age squared, gender.⁶⁷ When estimating the effect on political opinions I also control for feeling toward President Trump, worry about illegal immigration, does the participant believe Mueller is conducting a fair investigation, and whether the participant thinks Trump has attempted to

⁶⁶The time window used is not particularly important. If a 5-minute window is used instead the number of sites determined to have been visited through Facebook in the two weeks following the intervention decreases by less than 3%, and if a 60-minute window is used the number of sites increases by less than 3%.

⁶⁷Ideological leaning was not explicitly mentioned as a control variable in the pre-analysis plan. This covariate is included since it is used to determine if a participant was assigned to the pro-attitudinal or counter-attitudinal treatment. This does not affect the results.

obstruct the investigation into Russian interference in the election. When estimating the effect on affective polarization I control for the baseline value of the *feeling thermometer* measure and baseline value of the *difficult perspective* measure.

Age and gender are included in the Facebook data provided when participants log in to the survey and the remaining covariates are based on the baseline survey and defined in the following way

- Self-reported ideology is a nominal variable with seven ideological options from very liberal to very conservative and an option for participants who have not thought much about this.
- Party affiliation is a nominal variable with seven affiliation options ranging from strong Democrat to strong Republican along with an option of other party.
- Approval of Trump is a nominal variable with four options ranging from strongly disapprove to strongly approve. Ideological leaning is defined in Section 3.6.2.
- Feeling toward President Trump is an integer ranging from 0 to 100.
- Worry about illegal immigration is a nominal variable with the following options: not at all, only a little, fair amount, great deal.
- Whether Mueller is conducting a fair investigation and has Trump obstructed justice are both nominal variables with three options: yes, no, do not know.

In all regressions, if a covariate includes missing values, the missing values are coded to a constant and an additional dummy control is added to the regression indicating whether a value is missing. Regressions testing for heterogeneous effects also control for each participant's potential outlets since individuals who were assigned the alternative outlet may have different characteristics than individuals who were assigned the primary outlets.

C Pre-Analysis Plan

The main outcome and hypothesis tested in this study were pre-registered in the AEA RCT Registry.⁶⁸ The analysis deviates from the pre-analysis plan in two important ways. First, I use equal weights for the measures composing the indices, while in the plan I state that the weights for the index variables will be determined by the inverse of the covariance between variables at end-line (Anderson, 2008). This method is not used since it generates negative weights. When using negative weights the interpretation of the index is no longer clear. For example, the question on President Trump's approval rating received a negative weight according to this index, which means that *ceteris paribus*, a participant who has a more favorable opinion on Trump would be considered more liberal.

⁶⁸AEA RCT Registry Trial 0002713

Appendix Table A.17 repeats the analysis with the inverse-covariance weighting for affective polarization and political opinions. Column (1) of Appendix Table A.17b is the primary specification showing the effect of the counter-attitudinal treatment on affective polarization, compared to the pro-attitudinal treatment, using standard equal weights. Column (2) shows that the effect remains almost exactly the same when using inverse-covariance weights. One challenge with inverse-covariance weights is that this method does not cleanly generate weights for individuals with missing outcomes. In column (3), each outcome variable is first scaled, weights are then created using the inverse-covariance method based on participants with no missing outcomes, and the weights are renormalized to sum to one for each participant with missing outcomes, in order to create the index for all participants who have at least one non-missing outcome. The results remain essentially the same. Appendix Table A.17a shows the results of a similar analysis with the political opinion index. Since the inverse-covariance method generates negative weights, for some of the outcomes, columns (4) and (5) repeat the analysis with negative weights replaced with zero and the weights renormalized accordingly. While there is some variation in the results, the most straight-forward comparison is between columns (1) and (5). These columns focus on the same participants, and do not use different signs for the same weights, but assign different weights to the outcomes composing the index. In column (5) the effect of the conservative treatment is slightly larger, but still economically small and not statistically significant.

The second important deviation from the pre-analysis plan is that the polarization index originally included five attitudinal measures and three behavioral measures, while only the attitudinal measures are analyzed in this study. The behavioral measures were based on a question in the endline survey asking participants whether they would “like” or share a post stating that “In seeking truth, you have to get both sides of a story.” The primary behavioral outcome is composed of an index of the following measures: did participants state they will share the post, did participants state they would “like” the post, did participants actually share the post. However, it was not possible to analyze the posts of a large share of participants by the time they took the endline survey, partly due to the unexpected Cambridge Analytica scandal which led many individuals to revoke access to their posts. Furthermore, the behavioral measure turned out not to measure polarization well. While a measure of polarization should typically be correlated with partisanship, there was almost no correlation between being partisan and the behavioral outcomes.⁶⁹

Column (1) of Appendix Table A.18 shows that the primary estimate is still significant when using all eight variables in the polarization index. Column (3) measures the effect only on the behavioral outcomes (for most participants data not exist on whether posts were shared, therefore this index is mostly based on the self-reported survey answers). The effect of the treatments is small and not statistically significant. While this result does not change the conclusions regarding affective polarization, it is interesting to note that exposure to counter-attitudinal outlets does not affect individuals’ willingness to share a post regarding the importance of seeking both sides of a story.

⁶⁹The correlation between the behavioral polarization measures and partisanship is 0.03-0.06, while the correlation between the affective polarization measures and partisanship is 0.17-0.36.

D Additional Analysis

D.1 Heterogeneous Effects

In the pre-analysis plan, I stated that I will test for heterogeneous effects based on the whether participants are ideological, whether they are in an echo chamber, the openness of participants, and whether they are sophisticated.

I define participants as *Ideological* if the absolute value of their self-reported ideology on the 7 point scale (from -3 for very liberal to +3 for very conservative) is above or equals the median.

I define that participants are in an *Echo Chamber* if their answer to “Thinking about the opinions you see people post about government and politics on Facebook, how often are they in line with your own views” is above or equals median. I measure the *Potential Outlets Echo Chamber* based on whether the difference between how often participants reported seeing the potential pro-attitudinal and counter-outlets in their feed is above the median.⁷⁰

I measure whether a participant has an *Open Personality* according to whether her answer to the average of the following questions are above or below the median “I see myself as open to new experiences, complex” and the reverse values of “I see myself as conventional, uncreative”. The questions based on Gosling et al. 2003. I also measure a similar concept of whether the participants are *Certain* in their opinions based on the answer to “Generally speaking, how certain are you of your political opinions?” is above or equals the median.⁷¹

I define participants as *Sophisticated* if their answered one of the following questions correctly: “Suppose 110 members of a local government voted on an infrastructure bill. The bill passed by a margin of 100 votes. How many members voted against the bill”, “Suppose the number of US citizens on the internet doubles every month. If it took 48 months for the entire US population to have internet access, how many months did it take for half the population to have internet access”. These questions are based on the Cognitive Reflection Test (Shane, 2005).

In addition to the pre-registered tests, I explore the effect of several additional moderators. Social media use is measured using the following two variables. Participants are defined to get *Most News Social Media* if their answer to “Thinking specifically about government and politics, do you get most of your news about this topic...” is “Through social networking sites (such as Facebook or Twitter)”. Participants have *High News Subscriptions* if their baseline subscription to news outlets on Facebook is above or equals the median.

News engagement is measured using the following variables. Participants are considered *Exposed to Outlets* if their self-reported exposure to posts from the eight potential outlets in baseline is

⁷⁰In the pre-analysis plan I planned on using actual data from the extension to measure whether participants in an echo chamber, but data on the Facebook feed is not available before the treatment and post-treatment data may be biased.

⁷¹In the pre-analysis plan I planned on including a question on whether people changed their mind based on something that appeared in the past, however in order to keep the survey short, this question was eventually only asked in the endline survey, and I am using the certainty question instead.

above or equals the median. Participants are considered to *Know Outlets Slant* if the distance between their perceived slant of the potential outlets and the average perceived slant by participants with the same self-reported ideology is below the median. Participants are considered to *Follow the News* if their answer to "How often do you pay attention to what's going on in government and politics?" is above the median.

Finally, participants are considered *Conservative* if their ideological leaning is conservative and participants are considered *Older* if their age is above or equal to the median age.

The panel on the left side of Appendix Figure A.11 shows that the effect on political opinions is mostly homogeneous (i.e., most people were not persuaded by the treatment). In the panel, each row represents a separate regression estimating the effect of interacting the conservative treatment with the specified moderator, where the reference group is the liberal treatment. A higher value means that this group of participants were more likely to be persuaded by the treatment (they were more likely to become conservative, compared to other participants, when exposed to the conservative treatment, compared to the liberal treatment). Only two marginal significant effects are detected. Participants who self-report following the news more frequently were more likely to be affected by the treatment (the p-value is 0.11). Somewhat surprisingly, participants who scale higher on the openness index were *less* likely to be persuaded by the intervention.

The panel on the right side of Appendix Figure A.11 shows that there is no evidence for substantial heterogeneous effects on affective polarization, however, I am underpowered to detect mild heterogeneous effect. In the panel each row represents a separate regression estimating the effect of interacting the pro-attitudinal treatment with the specified variable, where the reference group is the counter-attitudinal treatment. A higher value means individuals were more likely to become polarized as a result of pro-attitudinal treatment, compared to the counter-attitudinal treatment. The strongest heterogeneous effect is between people who are more and less ideological. It seems that the treatment effects of the pro- and counter-attitudinal treatments are driven by individuals whose self-reported ideology is more moderate (respondent who define themselves as slightly liberal, moderate or slightly conservative). However, these differences are not statistically significant.

Appendix Figure A.12 presents the results from a specification estimating all heterogeneous effects in one regression for affective polarization and one regression for political opinions. The results are mostly similar.

D.2 Knowledge

While this paper focuses on persuasion and polarization, the survey included several questions related to political knowledge. The two primary measures of political knowledge are self-reported familiarity, whether participants reported hearing of news events and political figures, and accurate political knowledge, measured according to participants' answers to several true/false ques-

tions on recent events. For some questions, participants were expected to gain knowledge when assigned to the liberal outlet (heard of Michael Cohen, heard about the Stephon Clark shooting, believed the Russian government tried to influence the 2016 elections, believed a wall is not being built at the US-Mexico border) and for other measures, the conservative treatment was expected to have an effect (heard of Louis Farrakhan, heard about a controversial speech by Hillary Clinton in India, believed Trump is not a criminal target of the Mueller investigation, believed Trump's tax cuts will increase most people's income).

Table A.19 presents the effect of the treatment on knowledge for the four primary self-reported familiarity outcomes and the four primary accurate knowledge outcomes. The coefficients of interest are the effects of the liberal treatment on liberal outcomes and conservative treatment on conservative outcomes. The treatment seems to have little to no effect on the knowledge outcomes.⁷²

The browser extension data was used to test whether participants were exposed to differential coverage on topics included in the self-reported familiarity index. Table A.20 shows that the intervention affected news exposure. The regression measures the effect of the treatment on the number of posts mentioning each topic which appeared in the participants' social media feeds.⁷³

For all four topics, the treatment had a significant effect in the expected direction when the relevant treatment is compared to the control group, and for three of the four topics, the effect is also significant when the treatments are compared to each other.⁷⁴

The results presented in this section suggest that while the slant of one's social media news feed can determine what news events an individual is exposed to on social media, that exposure does not necessarily affect their political awareness of topics. One possible explanation is that individuals consume news also outside their social media feed. In any case, this result should not be interpreted as definitive evidence of a null effect. Participants were asked questions about very specific issues, the range of possible answers was limited, and answers to true/false questions could be driven by motivated reasoning and not by participants' true beliefs. Furthermore, previous studies have shown that the effect of media on political knowledge is complex, and depends on the context and the issue covered (Schroeder and Stone, 2015).

⁷²Besides one outcome (the effect on believing a wall is being constructed on the border), the effect are similar when running the regression only among participants who installed the extension for at least two weeks and completed the endline survey.

⁷³The number of posts discussing Michael Cohen, Louis Farrakhan or the shooting of Stephon Clark are measured by counting the number of posts mentioning the expression "Michael Cohen", "Louis Farrakhan" and "Stephon Clark", respectively. The fourth outcome is Hillary Clinton's speech in India suggesting that many white women voted for Trump since they took their voting cues from their husbands. To measure mentions of the speech I count the number of posts mentioning the words Clinton, vote and either India or husband.

⁷⁴The results are similar when running the regression only among participants who installed the extension for at least two weeks and completed the endline survey.

E Additional Figures and Tables

Figure A.1: Recruitment Ads

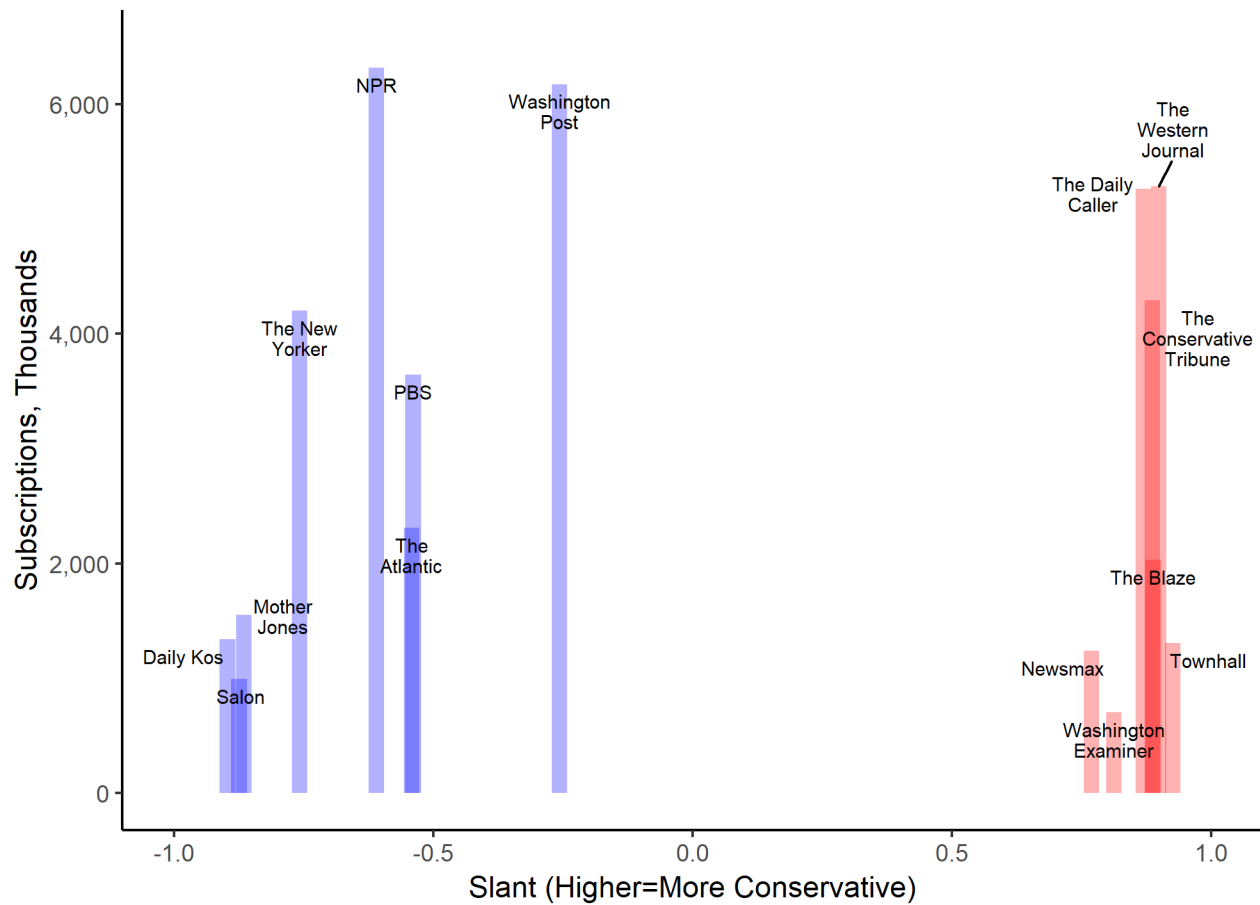


(a) Political Ad



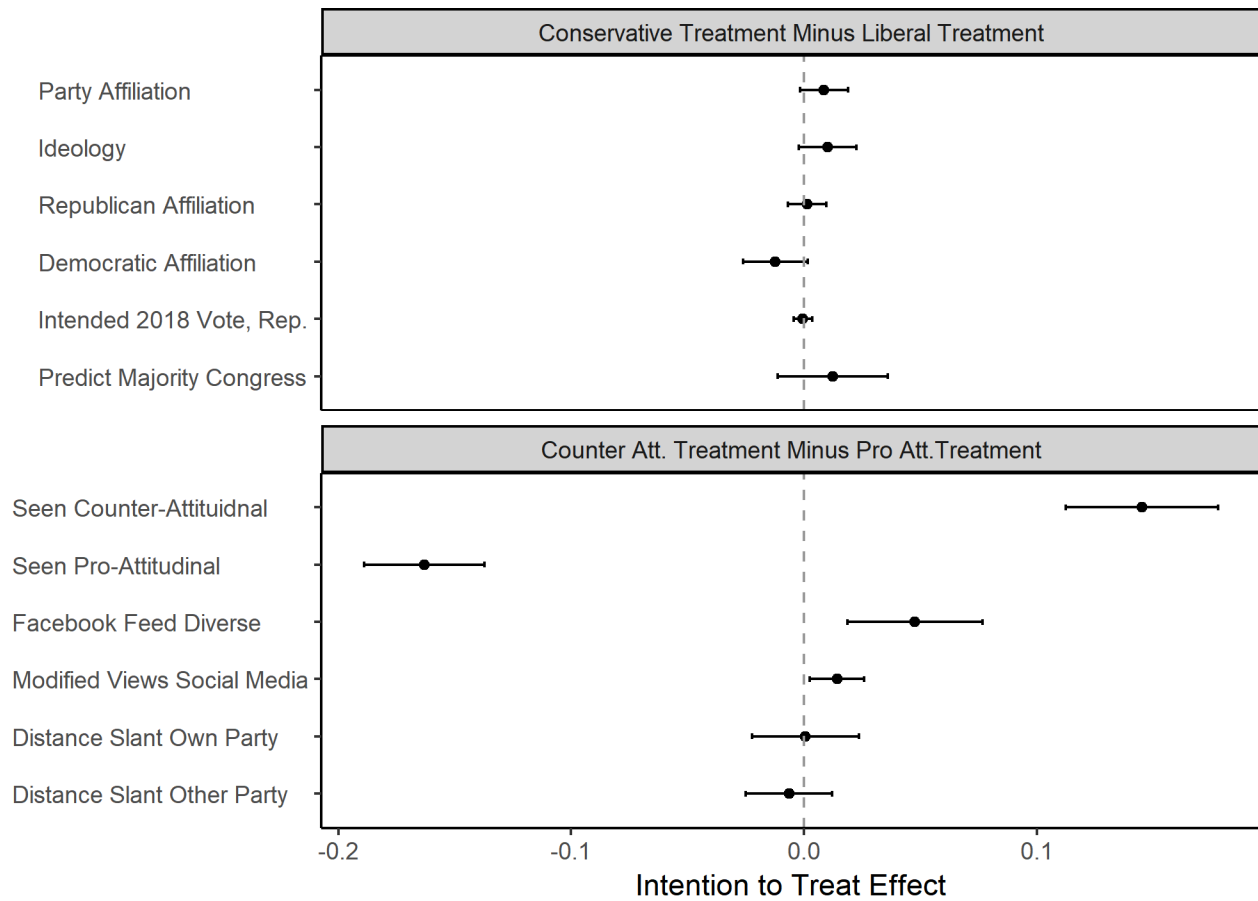
(b) General Ad

Figure A.2: Alternative Assigned Outlets



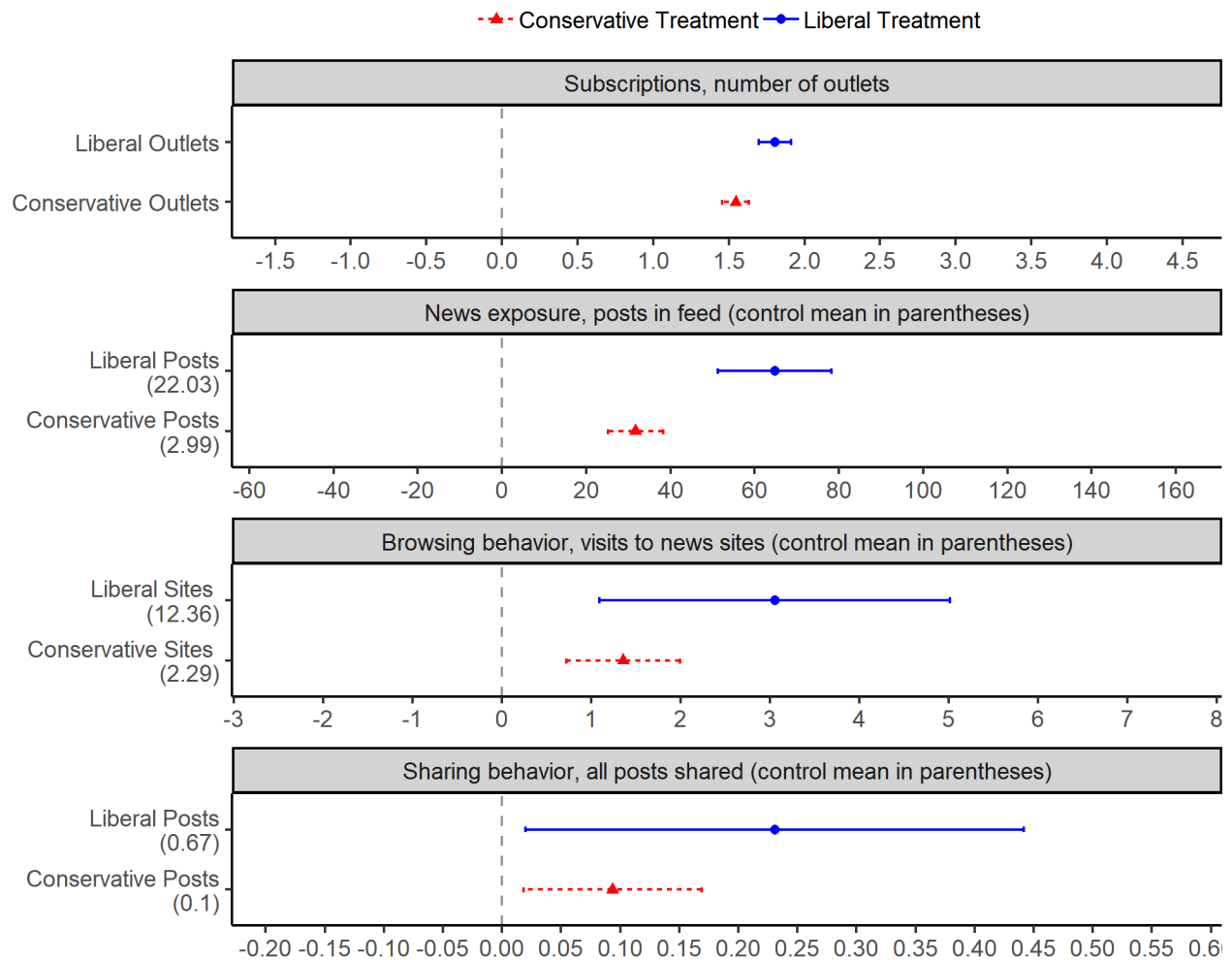
This figure displays the alternative outlets offered in the experiment if a participant already subscribed to one of the primary outlets. The x-axis is the slant of the outlets, as determined by Bakshy et al. (2015), and the y-axis is the total number of individuals who have subscribed to each outlet on Facebook in April 2018.

Figure A.3: Effect of the Treatment on Additional Outcomes



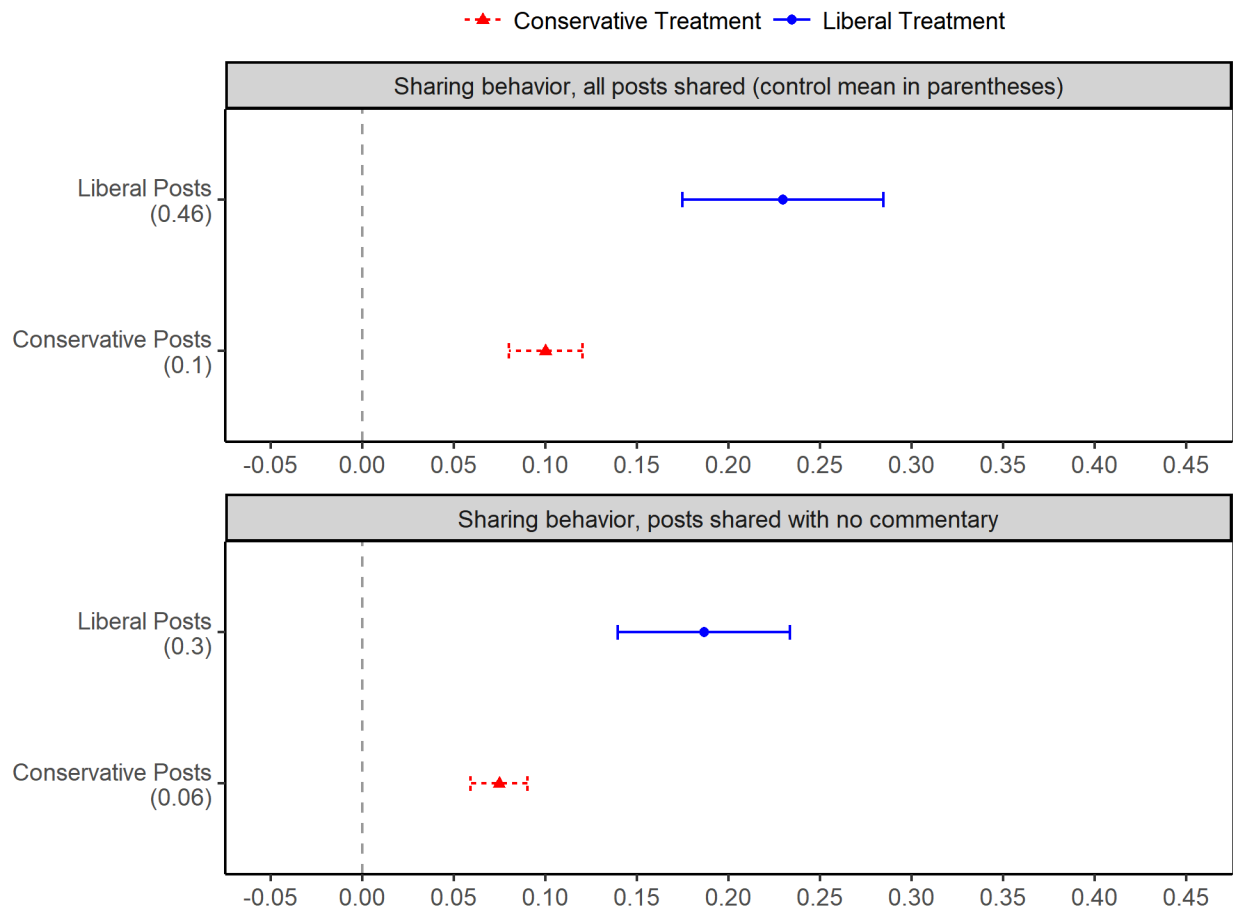
This figure shows the effect of the experiment on additional outcomes. *Party Affiliation* is the party the participant identifies with on a 7-point scale. *Ideology* is self-reported on a 7-point scale. *Republican/Democrat Affiliation* is coded as 1 if the participant leans toward the Republican/Democrat party, 2 if the participant is a Republican/Democrat, 3 if the participant is a strong Republican/Democrat and 0 otherwise. *Intended 2018 Vote, Rep.* is whether the participant would have voted for the Republican Party candidate in their district, if the election was held today, among participants intending to vote for the Republican or Democratic Party candidate. *Predict Majority Congress* is the party participants predicted will hold the majority of seats in Congress after the 2018 vote where the Republican Party is coded as 1, not sure is coded as 0 and the Democratic Party is coded as -1. *Seen Pro/Counter Attitudinal* is how often the participant reported seeing news from the pro or counter-attitudinal outlets in their Facebook feed over the past week where the possible answers are more than 5 times (3), 3-5 times (2), 1-2 times (1), Have not seen (0). *Facebook Feed Diverse* is the answer to the question “Thinking about the opinions you see about government and politics on Facebook, how often are they in line with your own views?”, where the possible answers are “Always or nearly all the time” (0), “Most of the time” (1), “Some of the time” (2), “Not too often” (3). *Modified Views Social Media* is whether consumers self-reported modifying their views in the past two months about a political or social issue because of something they saw on social media. *Distance Slant* is the difference between the participant’s baseline ideology and the perceived ideology of a party. Non-binary outcomes are standardized by subtracting the control group mean and dividing by the control group standard deviation. The specification and controls are described in more detail in Section 3.6. The Regressions also control for baseline measures of the outcomes when they exist. Error bars reflect 90 percent confidence intervals.

Figure A.4: Effects of the Liberal and Conservative Treatments on Subscriptions, News Exposure, News Sites Visited and Sharing Behavior, Two Weeks Following the Intervention



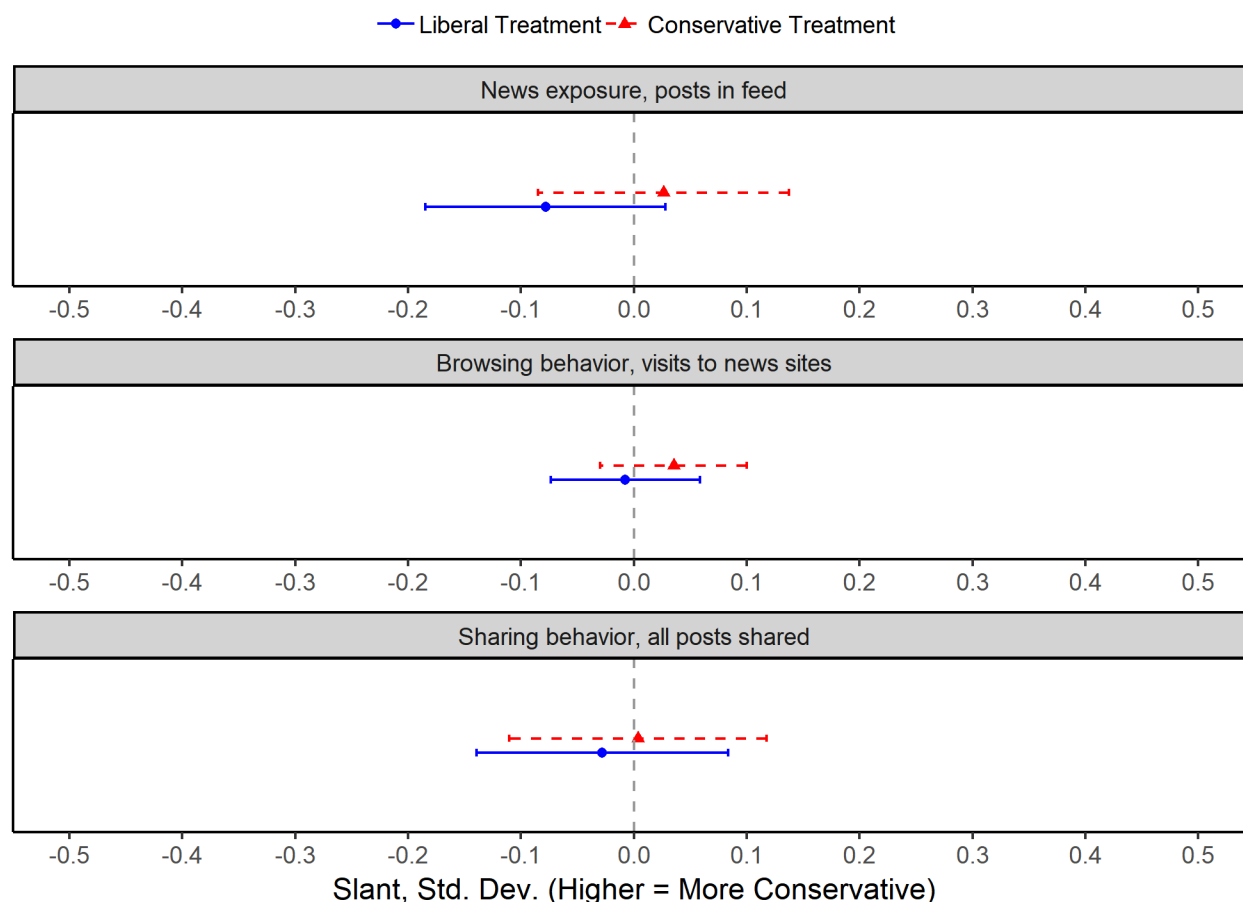
This figure shows the effect of the liberal and conservative treatments on engagement with each individual's potential outlets. Each row in the figure is estimated by regressing engagement with the four potential pro-attitudinal outlets or four potential counter-attitudinal outlets on the treatment. The outcomes are the number of outlets individuals subscribed to, the number of posts from the outlets that appeared in their feed, the number of times they visited the outlets' websites and the number of posts shared from the outlets. The figure includes data from 1,839 participants for which full data from the extension is available for the two weeks following the intervention. The regressions control for the outcome measure in baseline if it exists. Error bars reflect 90 percent confidence intervals.

Figure A.5: Effects of the Liberal and Conservative Treatments on Number of Posts Shared



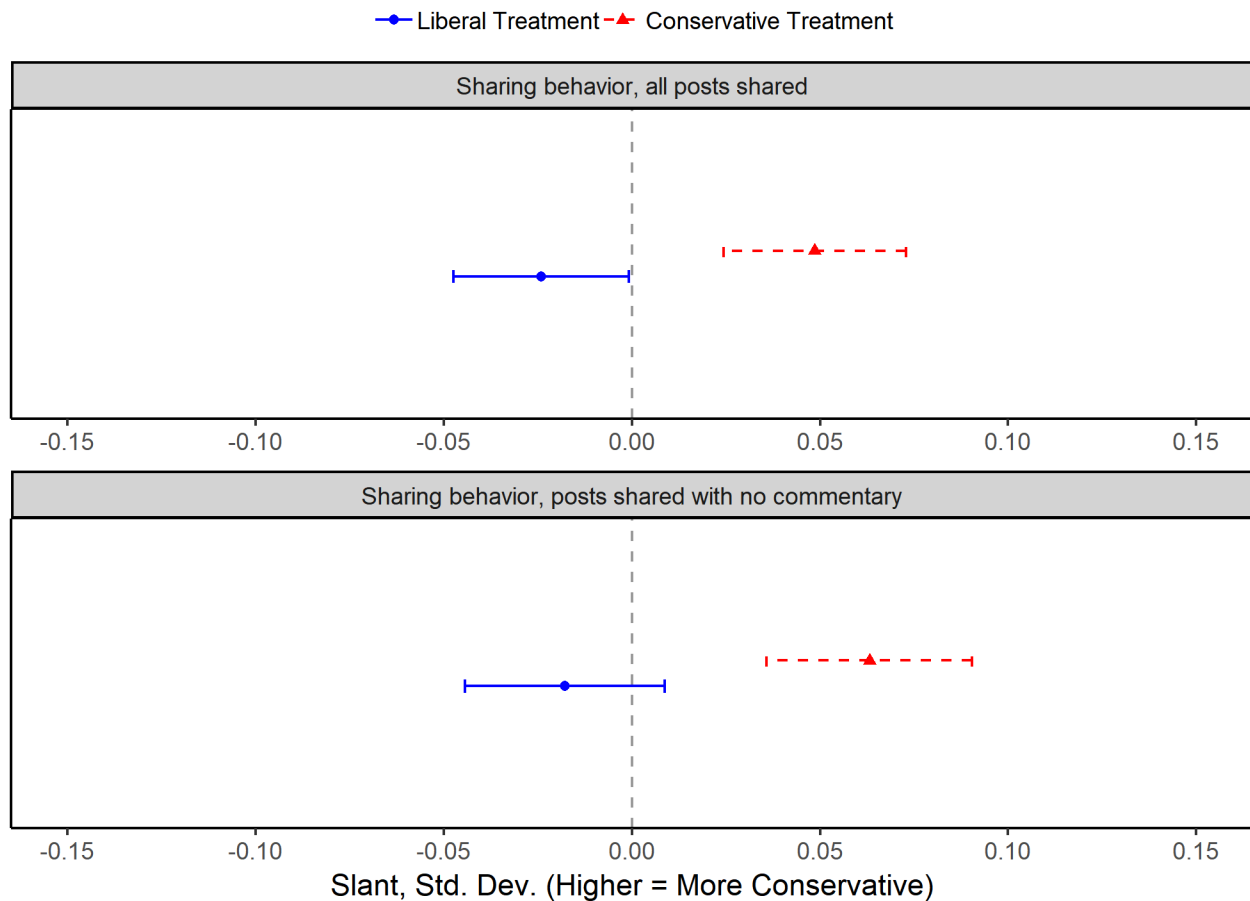
This figure shows the effect of the liberal and conservative treatments on engagement with each individual's four potential pro-attitudinal outlets and four potential counter-attitudinal outlets. Each row in the figure is estimated by regressing engagement with the four potential pro-attitudinal outlets or four potential counter-attitudinal outlets on the treatment. The outcomes in both panels are the number of posts the individuals shared from these outlets. The first panel includes all posts and the second panel includes only posts that were shared without any commentary by the participant. The figure includes data from 34,575 participants for which data on shared posts is available for at least two weeks following the intervention. The regressions control each outcome measure in baseline. Error bars reflect 90 percent confidence intervals.

Figure A.6: Effect of the Treatment on Slant of News Consumption, Excluding each Participant's Eight Potential Experimental Outlets



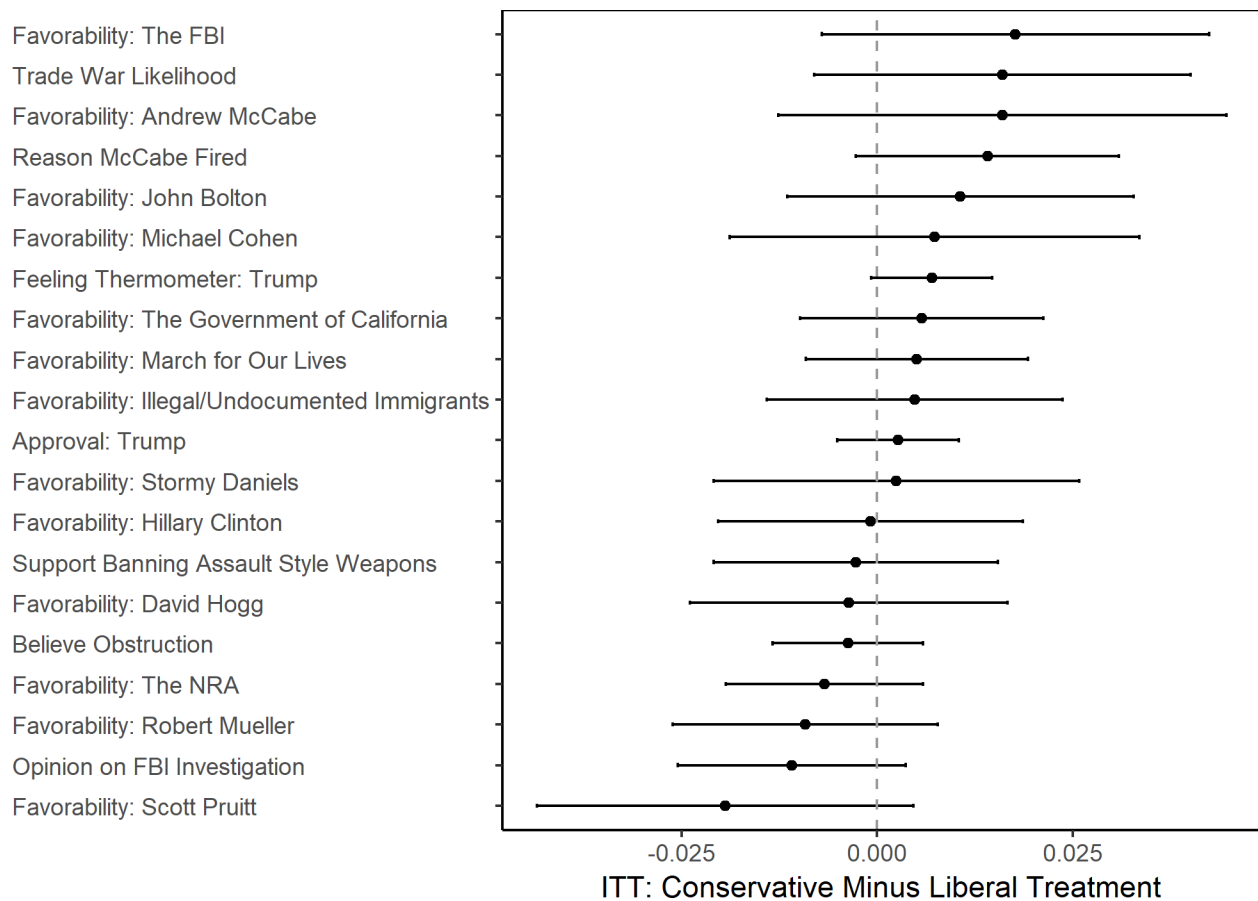
This figure shows the effect of the liberal and conservative treatments on the mean slant, in standard deviations, of all news individuals engaged with, excluding the four potential liberal outlets and the four potential conservative outlets defined for each participant. Each row in the figure is estimated by regressing engagement with the four potential conservative outlets or four potential liberal outlets on the treatment. The regressions control for the outcome in baseline if it exists. The figure displays the slant for three outcomes: exposure to posts on Facebook (panel 1), news sites visited (panel 2) and posts shared (panel 3). The first three outcomes include participants which installed the browser extension for at least two weeks, and the last outcome includes a larger sample of all participants who provided permissions to access their posts for at least two weeks. Error bars reflect 90 percent confidence intervals.

Figure A.7: Effects of the Liberal and Conservative Treatments on Slant of Posts Shared, Access Posts Sub-Sample



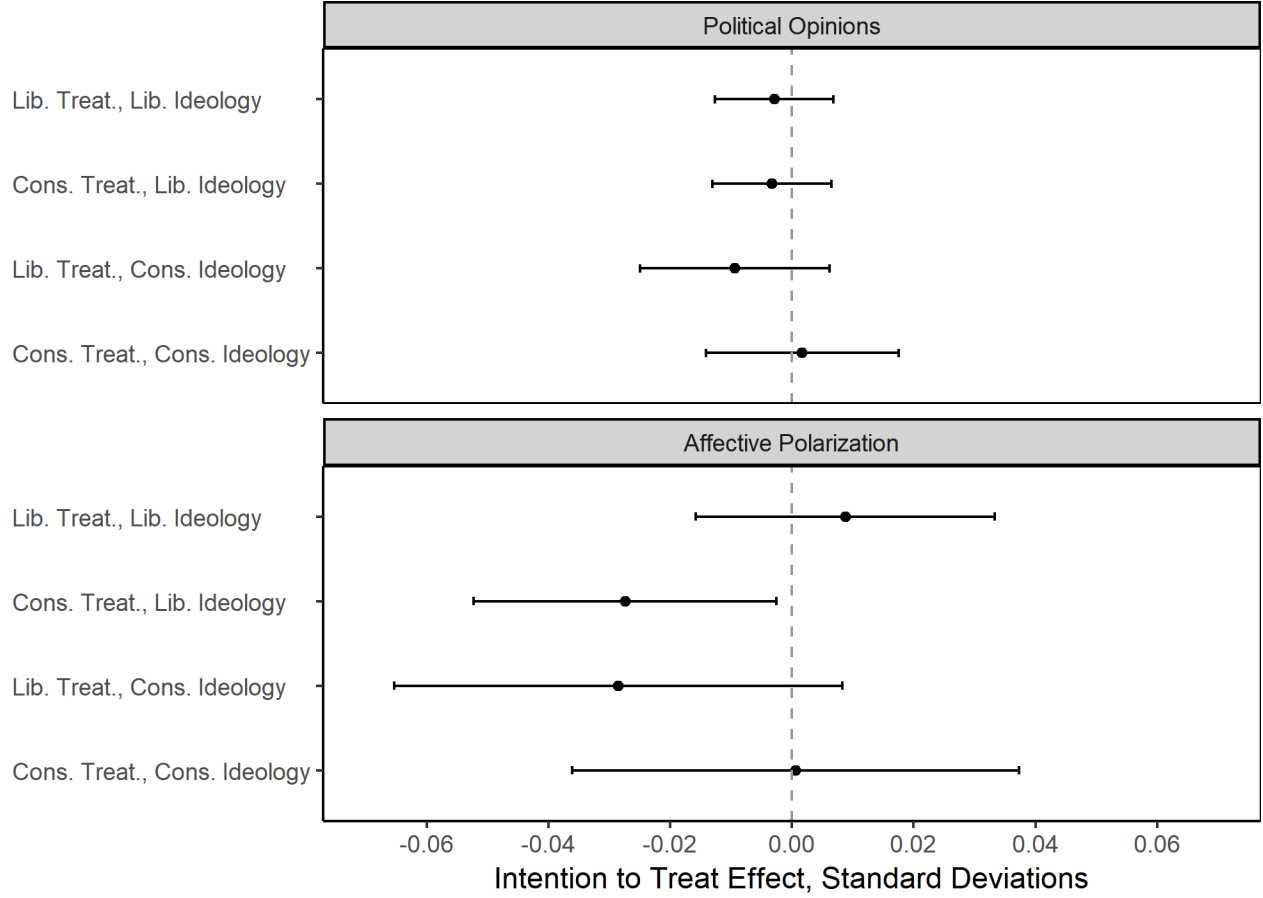
This figure shows the effect of the liberal and conservative treatments on the mean slant, in standard deviations, of all news individuals shared. Each panel in the figure is estimated by regressing the slant of posts shared on the treatment. The first panel includes all posts and the second panel includes only posts that were shared without any commentary by the participant. Data based on the access posts sub-sample: 34,575 participants who provided access to their posts for at least two weeks following the intervention. The regressions control each outcome measure in baseline. Error bars reflect 90 percent confidence intervals.

Figure A.8: Effect of the Treatment on Political Opinion Outcomes



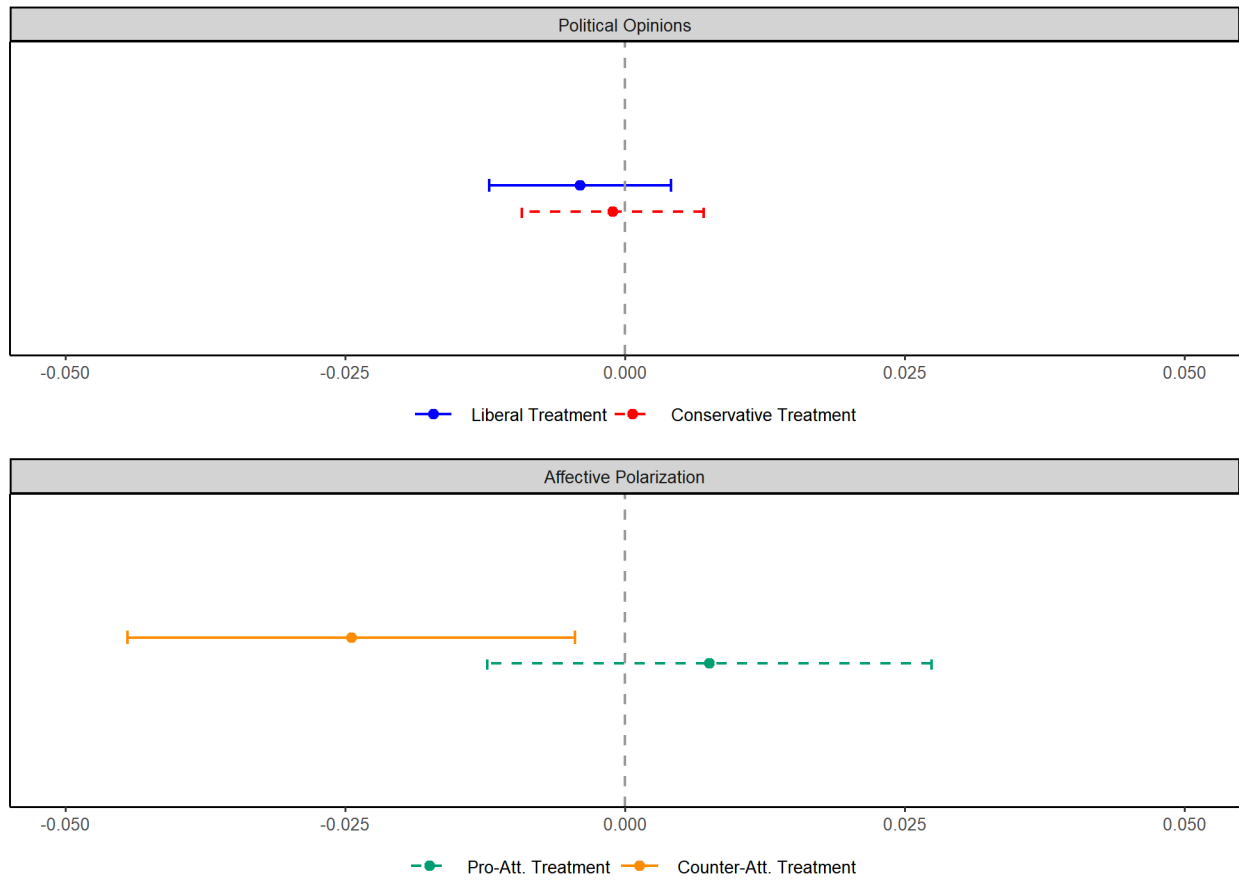
This figure shows the effect of the conservative treatment, compared to the liberal treatment on each outcome composing the political opinions index. Each row represents a separate regression as specified in Section 3.6. All the dependent variables have been standardized by subtracting each variable by the control group mean and dividing by the control group's standard deviation. Each variable is defined such that a higher value is associated with a more conservative outcome. The *Favorability* outcomes are based on questions asking participants whether they have a very favorable, favorable, unfavorable or very unfavorable opinion on specific individuals or organizations. *Approval: Trump* is whether participants strongly approve, somewhat approve, somewhat disapprove or strongly disapprove of the job Donald Trump is doing as President. *Feeling Thermometer: Trump* is how respondents feel towards trump on a 0-100 degrees scale. *Believe Obstruction* is whether participants believed that President Trump has attempted to derail or obstruct the investigation into the Russian interference in the 2016 election. *Opinion on FBI Investigation* is whether participants think the FBI investigation into Trump campaign officials' contacts with Russian government officials is a serious attempt to find out what really happened, a politically-motivated attempt to embarrass Donald Trump or equally-motivated by both of these. *Reason McCabe Fired* is whether participants believe McCabe was fired because of improper actions while serving as Deputy Director of the FBI, as a way to damage McCabe's credibility in any evidence he might give to the Robert Mueller investigation or as an act of revenge (multiple choice question). An answer of only improper action is coded as one, improper action along with damaging credibility or revenge is coded as 0, and an answer of only damaging credibility or revenge is coded as -1. *Trade War Likelihood* is whether participants believe it is very likely, somewhat likely, somewhat unlikely or very unlikely that a trade war will develop between the United States and foreign countries in the next year. *Support Banning Assault Style Weapons* is whether participants strongly support, support, oppose or strongly oppose banning assault-style weapons. Error bars reflect 90 percent confidence intervals.

Figure A.9: Effect of the Treatment on Political Opinions, by Baseline Ideology



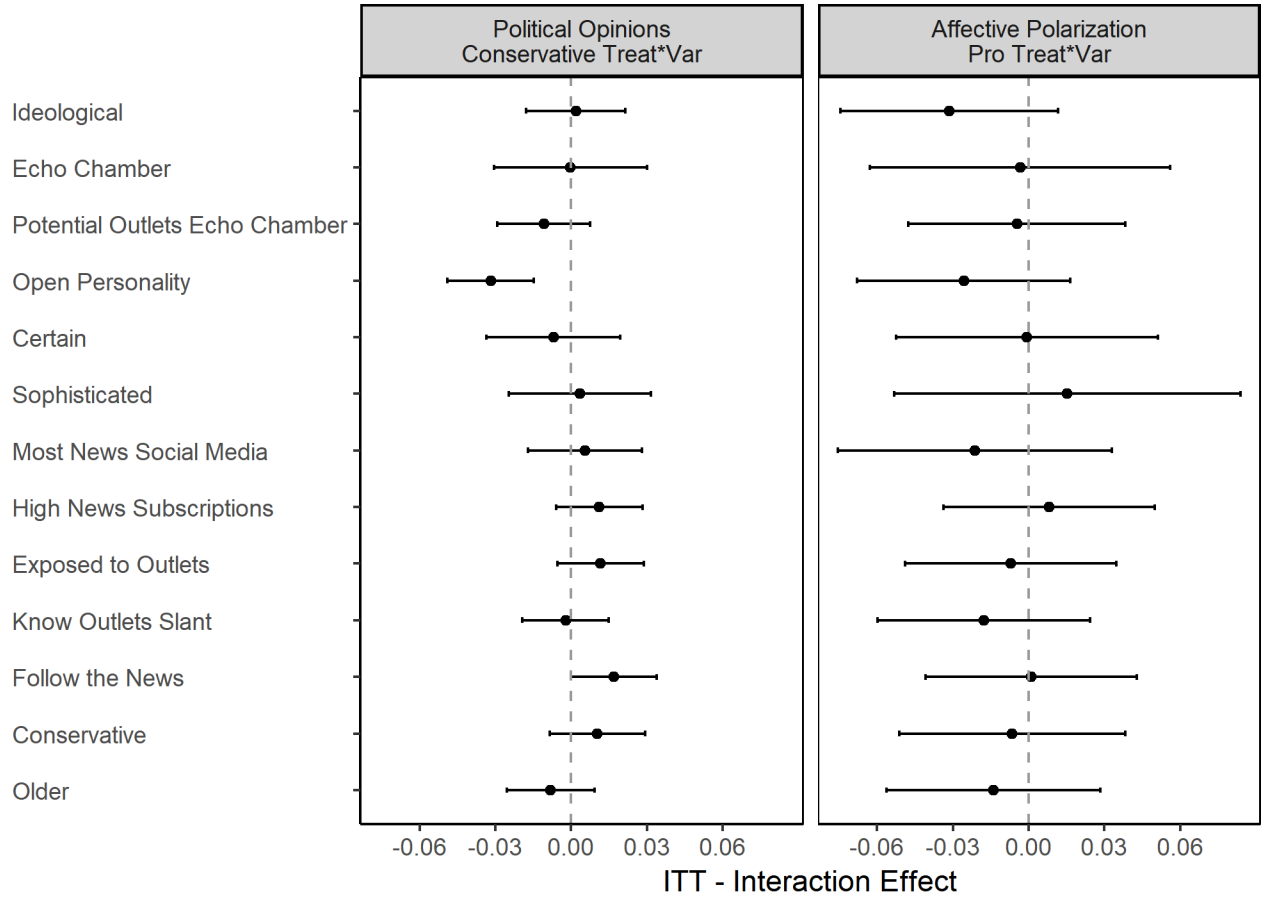
This figure shows the effect of the treatment among ideological subgroups based on the following model: $Y_i = \beta_1 T_i^L I_i^L + \beta_2 T_i^L I_i^C + \beta_3 T_i^C I_i^L + \beta_4 T_i^C I_i^C + I_i + \alpha X_i + \varepsilon_i$ where: T_i^C, T_i^L are binary indicators for the conservative and liberal treatments, I_i^C, I_i^L are binary indicators for whether the participants are conservative or liberal according to the baseline survey. The reference group is the control group. The controls are specified in Section 3.6. In the first panel, the x-axis is the intention to treat effect on the political opinions index, where a higher value is a more conservative outcome. In the second panel, the x-axis is the intention to treat effect on the affective polarization index, where a higher value is a more polarized outcome. Error bars reflect 90 percent confidence intervals.

Figure A.10: Effect of the Treatment by Treatment Arm



This figure shows the effect of each treatment arm on the political opinions index and affective polarization index. The indices are described in section 3.4.2. The specification and controls are described in more detail in Section 3.6. Error bars reflect 90 percent confidence intervals.

Figure A.11: Heterogeneous Effects on Political Opinions and Affective Polarization



In the Political Opinion Panel each row represents the β coefficient in the following separate regression:

$$Y_i = \alpha T_i^C + \beta T_i^C \times Var + \gamma Var + \delta X_i + \varepsilon_i,$$

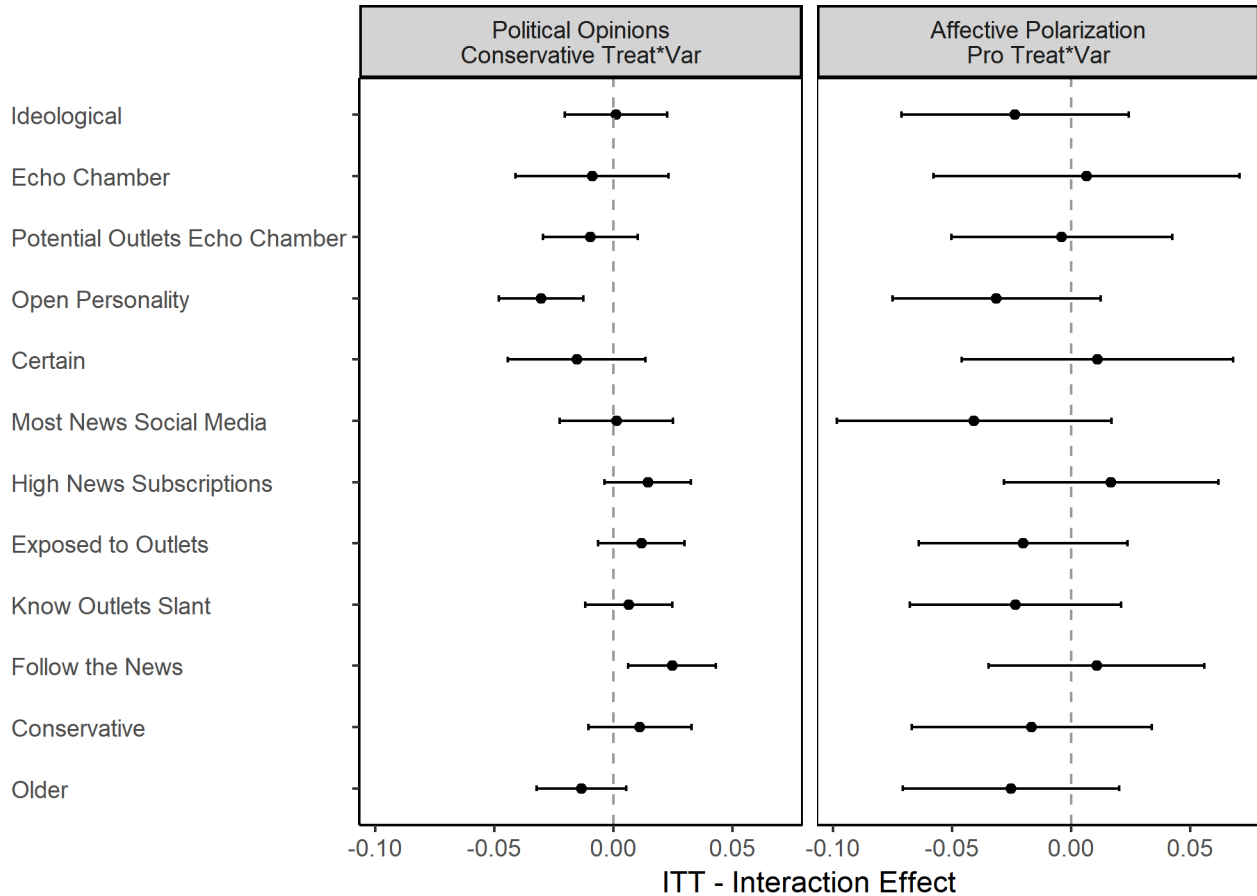
where the dependent variable is the political opinion index, the independent variable is the full interaction of the conservative treatment and the variable analyzed in the row and the control group is excluded so the reference category is the liberal treatment. A higher value means individuals were more likely to be persuaded by the treatment (they became more conservative as a result of the conservative treatment, compared to the liberal treatment).

In the Affective Polarization Panel each row presents the β coefficient in the following regression:

$$Y_i = \alpha T_i^P + \beta T_i^P \times Var + \gamma Var + \delta X_i + \varepsilon_i,$$

where the dependent variable is the affective polarization index, the independent variable is the full interaction of the pro-attitudinal treatment and the variable analyzed in the row and the control group is excluded so the reference category is the counter-attitudinal treatment. A higher value means individuals were more likely to become polarized as a result of pro-attitudinal treatment, compared to the counter-attitudinal treatment. The regressions control for the covariates specified in section 3.6. The definitions of the variables analyzed are described in section D.1. Error bars reflect 90 percent confidence intervals.

Figure A.12: Heterogeneous Effects on Political Opinions and Affective Polarization, Controlling for Other Heterogeneous effects



In the Political Opinion Panel each row represents the β coefficients in the following separate regression:

$$Y_i = \alpha T_i^C + \beta_1 T_i^C \times Var + \gamma_1 Var + \beta_2 T_i^C \times Var + \gamma_2 Var + \dots + \delta X_i + \varepsilon_i,$$

where the dependent variable is the political opinion index, the independent variable is the full interaction of the conservative treatment and all the variables displayed in the figure and the control group is excluded so the reference category is the liberal treatment. A higher value means individuals were more likely to be persuaded by the treatment (they became more conservative as a result of the conservative treatment, compared to the liberal treatment). This figure does not include the sophistication measure since many participants did not answer the questions composing the measure.

In the Affective Polarization Panel each row presents the β coefficients in the following regression:

$$Y_i = \alpha T_i^P + \beta_1 T_i^P \times Var + \gamma_1 Var + \beta_2 T_i^P \times Var + \gamma_2 Var + \dots + \delta X_i + \varepsilon_i,$$

where the dependent variable is the affective polarization index, the independent variable is the full interaction of the pro-attitudinal treatment and all the variables analyzed in the figure and the control group is excluded so the reference category is the counter-attitudinal treatment. A lower value means individuals were less likely to become polarized as a result of pro-attitudinal treatment, compared to the counter-attitudinal treatment. The regressions control for the covariates specified in section 3.6. The definitions of the variables analyzed are described in section D.1. Error bars reflect 90 percent confidence intervals.

Table A.1: Consuming News through Facebook and the Absolute Slant Consumed

	Absolute Slant	Absolute Slant	Mean Absolute Slant	Mean Absolute Slant	Mean Absolute Slant
	(1)	(2)	(3)	(4)	(5)
FB References	0.474*** (0.025)	0.335*** (0.013)			
FB Share of News Ref.			0.885*** (0.033)	0.496*** (0.018)	
FB Share of Visits					0.463*** (0.085)
Unit of Observation	Site	Site	Ind.	Ind.*Month	Ind.*Month
Individual FE		X		X	X
Month FE		X		X	X
Demographics			X		
Observations	2,260,860	2,260,860	59,707	272,236	272,236

This table shows the effect of social media on the absolute slant of news consumed. In columns (1)-(2) each observation is a website visited, *FB Reference* refers to a visit where the referring domain is "facebook.com" and the dependent variable is the absolute value of the slant of all sites visited, calculated based on Bakshy et al. (2015). In column (3) each observation is an individual and in columns (4)-(5) each observation is an individual*month. The dependent variable is measured in standard deviations. In columns (3)-(5) the dependent variable is the mean absolute value of the slant of all sites visited. *FB Share of News Ref* refers to the share of news sites an individual visited through Facebook and *FB Share of Visits* refers to the share of visits to Facebook among all websites visited. The sample includes all users who visited at least two news sites through Facebook and two news sites through other means. *p<0.1 **p<0.05 ***p<0.01

Table A.2: List of Outlets Offered and Subscriptions

Outlet	Group	Offered	Subscribed	Share
The Washington Times	Conservative	12365	3166	0.26
The National Review	Conservative	12056	2868	0.24
The Wall Street Journal	Conservative	11805	3914	0.33
Fox News	Conservative	10841	1387	0.13
The Daily Caller	Conservative	1470	309	0.21
The Western Journal	Conservative	509	146	0.29
Washington Examiner	Conservative	607	131	0.22
Townhall	Conservative	135	37	0.27
The Conservative Tribune	Conservative	72	31	0.43
The Blaze	Conservative	80	24	0.30
Newsmax	Conservative	32	13	0.41
Slate	Liberal	11737	2938	0.25
MSNBC	Liberal	11687	2716	0.23
HuffPost	Liberal	10642	2304	0.22
The New York Times	Liberal	10143	3285	0.32
Washington Post	Liberal	2825	1296	0.46
Salon	Liberal	1668	572	0.34
Daily Kos	Liberal	661	222	0.34
NPR	Liberal	119	66	0.55
Mother Jones	Liberal	150	58	0.39
The Atlantic	Liberal	203	111	0.55
The New Yorker	Liberal	105	65	0.62
PBS	Liberal	40	23	0.57

This table shows the list of all outlets included in the experiment. *Offered* is the number of participants in the baseline sample who were offered to subscribe to the outlet. *Subscribed* is the number of participants who subscribed to each outlet. *Share* is subscribed divided by offered.

Table A.3: Descriptive Statistics by Sample

	Baseline Sample	Access-Posts Sub-sample	Endline Survey Sub-sample	Browser Extension Sub-sample
Ideology (-3, 3)	-0.61	-0.61	-0.70	-0.95
Ideology, Abs. Value (0, 3)	1.75	1.75	1.80	1.81
Republican	0.17	0.17	0.16	0.14
Democrat	0.38	0.38	0.40	0.44
Feeling Therm, Rep.	29.1	29.2	27.5	22.8
Feeling Therm, Dem.	47.0	47.0	47.8	51.2
Feeling Therm, Difference	50.2	50.3	50.3	51.1
Difficult Pers., Difference	1.92	1.92	1.96	1.92
Most News Social	0.18	0.18	0.17	0.16
Took Survey Mobile	0.67	0.67	0.63	0.00
Female	0.52	0.52	0.52	0.49
Age	48.5	48.4	49.6	53.4
Total Subscriptions	474	474	472	477
News Outlets Subscriptions	8.47	8.48	8.65	9.00
Compliance	0.50	0.51	0.57	0.76
N	11,769	34,575	5,564	1,839

This table presents descriptive statistics by sub-sample. Baseline sample includes all participants, *Access-posts sub-sample* include participants who provided access to posts they shared for at least two weeks. *Endline survey sub-sample* include all participants who completed the baseline survey. *Browser extension sub-sample* includes all participants who installed the browser extension for at least two weeks. *Abs Ideology* is the absolute value of self-reported ideology. *Feeling Therm, Difference* is the difference between the feeling toward the participant's party and the opposing party according to the feeling thermometer questions. *Difficult Pers., Difference* is the difference in whether participants find it difficult to see things from the opposing party and their own party. For all other variables see Table 3.

Table A.4: Balance Table by Assignment to the Pro-attitudinal and Counter-attitudinal Treatments

Variable	Mean		Difference Between Treatments		
	Sample	US	Control - Pro.	Control - Counter.	Pro. - Counter.
Baseline Survey					
Ideology, Abs. Value (0, 3)	1.75	1.22	0.002	-0.000	-0.002
Republican	0.17	0.25	0.001	-0.006	-0.007
Independent	0.37	0.29	-0.010*	0.002	0.012**
Democrat	0.38	0.3	0.009	0.004	-0.005
Vote Support Clinton	0.53		-0.004	-0.003	0.001
Vote Support Trump	0.26		0.000	0.004	0.004
Feeling Therm, Difference	50.2	39.2	0.351	0.399	0.048
Difficult Pers., Difference	1.92		0.031	0.015	-0.016
Most News Social	0.18	0.13	0.004	-0.004	-0.008*
Device					
Took Survey Mobile	0.67		-0.011*	-0.001	0.010*
Facebook					
Female	0.52	0.51	-0.006	-0.003	0.003
Age	48.5	47.5	-0.013	0.048	0.061
Total Subscriptions	474		8.181	2.423	-5.758
News Outlets Subscriptions	8.47		-0.052	-0.086	-0.034
News Outlets Slant, Abs. Value	0.55		-0.004	-0.005*	-0.001
Access Posts, Pre-Treatment	0.98		0.002	0.002	-0.000
Attrition					
Took Followup Survey	0.47		0.027***	0.029***	0.003
Access Posts, 2 Weeks	0.92		0.006	0.002	-0.003
Ext Install, 2 Weeks	0.05		0.002	-0.000	-0.002
F-Test			1.250	0.794	1.087
P-value			[0.189]	[0.743]	[0.351]

This table presents descriptive statistics by whether participants were assigned to the pro-attitudinal treatment, counter-attitudinal treatment or control group. *Ideology, Abs. Value* is the absolute value of self-reported ideology. *Feeling Therm, Difference* is the difference between the feeling toward the participant's party and the opposing party according to the feeling thermometer questions. *Difficult Pers., Difference* is the difference in whether participants find it difficult to see things from the opposing party and their own party. *News Outlets Slant, Abs. Value* is the absolute value of the mean slant of all outlets participants subscribed to on Facebook in baseline. Slant range from -1 to 1 and is based on Bakshy et al. (2015) For all other variables see Table 3. *p<0.1 **p<0.05 ***p<0.01

Table A.5: Balance Table by Assignment to the Liberal and Conservative Treatments, Among Participants who Completed the Follow-up Survey

Variable	Mean		Difference Between Treatments		
	Sample	US	Control - Lib.	Control - Cons.	Cons. - Lib.
Baseline Survey					
Ideology (-3, 3)	-0.70	0.15	-0.010	-0.023	0.013
Republican	0.16	0.25	0.001	0.001	0.001
Independent	0.36	0.29	-0.016*	-0.007	-0.009
Democrat	0.40	0.3	0.014	0.008	0.006
Vote Support Clinton	0.55		-0.002	-0.000	-0.002
Vote Support Trump	0.25		0.009	-0.002	0.012
Feeling Therm, Rep.	27.5	43.1	0.224	-0.051	0.275
Feeling Therm, Dem.	47.8	48.7	0.400	0.684	-0.284
Difficult Pers., Rep. (1, 5)	3.18		0.040	0.005	0.035
Difficult Pers., Dem. (1, 5)	2.35		-0.006	-0.034	0.028
Most News Social	0.17	0.13	-0.013*	0.000	-0.014*
Device					
Took Survey Mobile	0.63		-0.007	0.007	-0.015
Facebook					
Female	0.52	0.51	-0.006	-0.004	-0.003
Age	49.6	47.5	0.489	-0.425	0.914**
Total Subscriptions	472		3.084	15.094	-12.011
News Outlets Subscriptions	8.65		-0.027	-0.088	0.061
News Outlets Slant (-1, 1)	-0.22		0.001	-0.005	0.006
Access Posts, Pre-Treatment	0.98		0.000	0.004	-0.004
F-Test			1.270	0.743	1.160
P-value			[0.196]	[0.769]	[0.286]

This table presents descriptive statistics by whether participants were assigned to the liberal treatment, conservative treatment or control group among participants who completed the endline survey. The variables are explained in the notes for Table 3. *p<0.1 **p<0.05 ***p<0.01

Table A.6: Balance Table by Assignment to the Pro-attitudinal and Counter-attitudinal Treatment, among Participants who Completed the Follow-up Survey

Variable	Mean		Difference Between Treatments		
	Sample	US	Control - Pro.	Control - Counter.	Pro. - Counter.
Baseline Survey					
Ideology, Abs. Value (0, 3)	1.80	1.22	-0.001	0.003	0.004
Republican	0.16	0.25	0.001	0.001	-0.000
Independent	0.36	0.29	-0.018**	-0.004	0.014
Democrat	0.40	0.3	0.015	0.008	-0.007
Vote Support Clinton	0.55		-0.002	-0.000	0.002
Vote Support Trump	0.25		0.001	0.007	0.006
Feeling Therm, Difference	50.3	39.2	0.925*	1.103**	0.178
Difficult Pers., Difference	1.96		0.050	0.038	-0.012
Most News Social	0.17	0.13	-0.004	-0.009	-0.005
Device					
Took Survey Mobile	0.63		-0.007	0.007	0.014
Facebook					
Female	0.52	0.51	-0.007	-0.007	0.000
Age	49.6	47.5	0.086	0.059	-0.027
Total Subscriptions	472		6.740	2.263	-4.476
News Outlets Subscriptions	8.65		-0.036	-0.120	-0.083
News Outlets Slant, Abs. Value	0.55		-0.001	0.001	0.002
Access Posts, Pre-Treatment	0.98		-0.000	0.004	0.004
F-Test			0.538	0.779	0.545
P-value			[0.942]	[0.728]	[0.938]

This table presents descriptive statistics by whether participants were assigned to the pro-attitudinal treatment, counter-attitudinal treatment or control group among participants who completed the endline survey. The variables are explained in the notes for Tables 3 and A.4. *p<0.1
 p<0.05 *p<0.01

Table A.7: Descriptive Statistics by Compliance

	Control	Pro-Attitudinal			Counter-Attitudinal			Liberal			Conservative		
		Comply	Non Comply		Comply	Non Comply		Comply	Non Comply		Comply	Non Comply	
Ideology (-3, 3)	-0.62	-0.87	-0.34	-1.04	-0.29	-1.13	-0.10	-0.72	-0.52				
Ideology, Abs. Value (0, 3)	1.80	1.84	1.75	1.78	1.82	1.78	1.72	1.75	1.75				
Republican	0.17	0.15	0.20	0.13	0.22	0.11	0.24	0.16	0.18				
Independent	0.35	0.35	0.38	0.36	0.35	0.36	0.38	0.37	0.36				
Democrat	0.40	0.44	0.33	0.45	0.35	0.47	0.29	0.40	0.37				
Vote Support Clinton	0.54	0.60	0.48	0.63	0.47	0.64	0.41	0.55	0.50				
Vote Support Trump	0.27	0.23	0.33	0.17	0.35	0.15	0.37	0.25	0.28				
Feeling Therm, Difference	50.5	51.0	49.0	48.9	51.0	50.4	49.7	49.7	50.5				
Difficult Pers., Difference	1.93	1.97	1.82	1.88	1.95	1.94	1.89	1.92	1.88				
Most News Social	0.17	0.17	0.17	0.19	0.17	0.18	0.17	0.18	0.17				
Took Survey Mobile	0.67	0.66	0.70	0.66	0.67	0.67	0.68	0.65	0.69				
Female	0.52	0.56	0.48	0.59	0.46	0.58	0.46	0.56	0.47				
Age	48.7	49.4	47.9	48.2	49.0	48.5	48.1	49.0	48.5				
Total Subscriptions	476	496	433	526	431	516	431	507	432				
News Outlets Subscriptions	8.52	9.20	7.79	9.17	8.14	9.18	7.81	9.05	7.92				
Echo Chamber	1.20	1.23	1.15	1.22	1.19	1.23	1.14	1.19	1.17				
N	12,105	6,741	5,359	5,505	6,617	6,279	6,216	6,312	6,181				

This table presents descriptive statistics on compliance by treatment arm for the entire baseline sample. For an explanation on each variable see Table 3.

Table A.8: Effect on Slant by Subsample

	News Exposure			Browsing Behavior			Shared Posts		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Liberal Treatment	-0.251*** (0.065)	-0.248*** (0.068)	-0.212*** (0.080)	-0.093** (0.039)	-0.082** (0.041)	-0.103** (0.048)	-0.024* (0.015)	-0.111 (0.068)	-0.071 (0.079)
Conservative Treatment	0.345*** (0.064)	0.355*** (0.067)	0.450*** (0.079)	0.098** (0.039)	0.101** (0.040)	0.103** (0.048)	0.049*** (0.015)	0.064 (0.067)	0.115 (0.079)
Cons. Treat. - Lib. Treat.	0.60*** (0.06)	0.60*** (0.07)	0.66*** (0.08)	0.19*** (0.04)	0.18*** (0.04)	0.21*** (0.05)	0.07*** (0.01)	0.17** (0.07)	0.19** (0.08)
Ext Sub-sample	X			X			X		
Posts Sub-sample		X			X			X	
Ext + Posts Sub-sample			X			X			X
Ext + Posts + Endline Sub-sample									
Observations	1,557	1,434	1,011	1,788	1,655	1,168	14,336	803	611

This table presents descriptive statistics on the effect of slant by the sub-sample analyzed. The dependent variables are the mean slant in standard deviations of news participants were exposed to in their feed (column 1-3), of news sites they visited (columns 4-6) and of news they shared (columns 7-9). *Ext Sub-sample* refers to the extension sub-sample, i.e. all participants who installed the extension for at least two weeks. *Posts Sub-sample* refers to the access posts sub-sample, i.e. all participants who provide permissions to access their posts for at least two weeks. *Ext + Posts Sub-sample* refers to all participants in both these sub-samples. *Ext + Posts + All Sub-sample* refers to all participants who installed the extension, provided access to posts and completed the endline survey. Regression control for outcome variables in baseline when they exist. Robust standard errors. *p<0.1 **p<0.05 ***p<0.01

Table A.9: Effect of Conservative and Liberals Switching Facebook Feeds

FB Slant, Std. Dev.	0.040 (0.036)
First Stage F	50.77
Control Difference in Slant: Conservative - Liberal	1.574
Effect of Switching Feeds	0.063
Control Difference in Pol. Opinoin: Conservative - Liberal	1.878
Effect of Switching Feed, Share of Control Group	3%
Observations	1,123

The table shows how the political opinions index would have changed if the slant of posts in the Facebook feed of a liberal participant would have been similar to the posts in the feeds of conservative (or vice versa). The calculation is based on the following steps. *FB Slant, Std. Dev.* shows the effect of the slant of posts in the Facebook feed on the political opinion index. The regressions are IV regressions with the treatment as the instrument and controlling for the covariates specified in section 3.6. *Control Difference in Slant* is the difference in the slant of posts in the Facebook feed of liberals and conservatives in the control group. *Effect of Switching Feeds* multiplies the effect found with the difference to show how much opinions would have changed if the feeds of liberals and conservatives had the same slant. *Control Difference in Slant* compares this effect to the actual difference between the opinions of conservatives and liberals. *Effect of Switching Feed, Share of Control Group* divides the effect of switching feed by the difference in the control group to calculate the final figure discussed in section 5.1. The sample is all posts between the baseline and endline surveys for participants who installed the extension for at least two weeks. Robust standard errors. *p<0.1 **p<0.05 ***p<0.01

Table A.10: Primary Outcomes, Controlling for Covariates

(a) Effect of the Treatment on the Political Opinions Index

	(1)	(2)	(3)
Conservative Treatment	0.011 (0.018)	0.001 (0.006)	-0.001 (0.005)
Liberal Treatment	-0.007 (0.018)	-0.005 (0.006)	-0.004 (0.005)
Cons. Treatment - Lib. Treatment	0.018 (0.019)	0.006 (0.006)	0.003 (0.005)
Common Controls		X	X
Baseline Political Opinions Controls			X
Observations	17,634	17,128	17,128

(b) Effect of the Treatment on the Affective Polarization Index

	(1)	(2)	(3)
Pro-Att. Treatment	-0.020 (0.019)	-0.001 (0.015)	0.008 (0.012)
Counter-Att. Treatment	-0.052*** (0.019)	-0.036** (0.015)	-0.024** (0.012)
Pro-Att. Treat. - Counter-Att. Treatment	0.032* (0.019)	0.035** (0.015)	0.032*** (0.012)
Common Controls		X	X
Baseline Affective Polarization Controls			X
Observations	16,894	16,894	16,894

These tables present the effect of the treatments on the political opinions index and the affective polarization index. Column (1) does not control for any covariates. Column (2) controls for self-reported ideology, party affiliation, 2016 candidate supported, ideological leaning, age, age squared, and gender. Column (3) also controls for baseline question similar to endline questions composing each index. The specification and controls are described in more detail in Section 3.6. Robust standard errors. * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

Table A.11: Primary Outcomes, According to Outlets Offered

(a) Effect of the Treatment on the Political Opinions Index

	(1)	(2)	(3)
Liberal Treatment	−0.004 (0.005)	−0.007 (0.007)	−0.005 (0.005)
Conservative Treatment	−0.001 (0.005)	−0.006 (0.007)	−0.002 (0.005)
Cons. Treat - Lib. Treat	0.003 (0.005)	0.001 (0.007)	0.003 (0.005)
Controls	X	X	X
Include Participants Who Already Subscribed To A Primary Outlet In Baseline	X		X
Potential Outlets FE			X
Observations	17,128	9,259	17,128

(b) Effect of the Treatment on the Affective Polarization Index

	(1)	(2)	(3)
Pro-Att. Treatment	0.008 (0.012)	0.002 (0.016)	0.006 (0.012)
Counter-Att. Treatment	−0.024** (0.012)	−0.026 (0.016)	−0.028** (0.013)
Pro-Att. Treat. - Counter-Att. Treat	0.032*** (0.012)	0.029* (0.016)	0.034*** (0.013)
Controls	X	X	X
Include Participants Who Already Subscribed To A Primary Outlet In Baseline	X		X
Potential Outlets FE			X
Observations	16,894	9,129	16,894

These tables present the effect of the treatments on the political opinions index and the affective polarization index. Column (1) is the primary specification and includes all participants. Column (2) includes only participants who did not subscribe in baseline to any of the four primary liberal outlets or the four primary conservative outlets. Thus, in this column, all participants in the liberal treatment were offered the same four primary liberal outlets and all participants in the conservative treatment were offered the same conservative outlets. Column (3) controls for the set of eight potential liberal and conservative outlets defined for each participant. The specification and controls are described in more detail in Section 3.6. Robust standard errors. * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

Table A.12: Primary Outcomes, by Sub-Sample

(a) Effect of the Treatment on the Political Opinions Index

	(1)	(2)	(3)	(4)
Liberal Treatment	−0.004 (0.005)	−0.005 (0.005)	−0.010 (0.018)	−0.018 (0.019)
Conservative Treatment	−0.001 (0.005)	−0.003 (0.005)	0.005 (0.018)	0.002 (0.018)
Cons. Treat - Lib. Treat	0.003 (0.005)	0.003 (0.005)	0.015 (0.018)	0.020 (0.018)
Controls	X	X	X	X
Sample	Endline	Endline+Posts	Endline+Ext	Endline+Posts+Ext
Observations	17,128	15,862	1,253	1,163

(b) Effect of the Treatment on the Affective Polarization Index

	(1)	(2)	(3)	(4)
Pro-Att. Treatment	0.008 (0.012)	0.010 (0.013)	0.025 (0.043)	0.036 (0.045)
Counter-Att. Treatment	−0.024** (0.012)	−0.024* (0.013)	−0.070 (0.042)	−0.053 (0.044)
Pro-Att. Treat. - Counter-Att. Treat	0.032*** (0.012)	0.034*** (0.013)	0.095** (0.043)	0.089** (0.044)
Controls	X	X	X	X
Sample	Endline	Endline+Posts	Endline+Ext	Endline+Posts+Ext
Observations	16,894	15,639	1,243	1,153

These tables present the effect of the treatments on the political opinions index and the affective polarization index. Column (1) is the primary specification and includes all participants who completed the endline survey (the endline survey sub-sample). Column (2) includes only participants who also provided participants to access their posts for at least two weeks (participants in the endline survey sub-sample and the access posts sub-sample). Column (3) includes only participants who installed the extension for at least two weeks (participants in the endline survey sub-sample and the extension sub-sample). Column (4) includes only participants who both provided access to posts and installed the extension (participants in all sub-samples). The specification and controls are described in more detail in Section 3.6. Robust standard errors. * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

Table A.13: Effect of a Balanced Facebook Feed

	Affective Polarization: Std. Dev.	Feeling Thermometer: Degrees
	(1)	(2)
FB Counter-Att. Share	-0.591** (0.284)	-11.413 (7.677)
Control: Counter-Att. Share	0.169	0.169
Effect of Balanced Facebook Feed	-0.195	-3.762
Observations	1,071	1,030

The table shows how affective polarization would have changed if the Facebook feed was balanced. The calculation is based on the following steps. *FB Counter-Att. Share* shows the effect of the share of counter-attitudinal news on the affective polarization index and the feeling thermometer measure. The regressions are IV regression with the treatment as the instrument and controlling for the covariates specified in section 3.6. *Control: Counter-Att. Share* shows that in the control group, approximately 17% of posts were counter-attitudinal. *Effect of Balanced Facebook Feed* shows how affective polarization and the feeling thermometer outcome would have decreased if the share of counter-attitudinal posts increased to 50%. Data includes all posts participants in the extension sub-sample were exposed to between the baseline and endline surveys. Robust standard errors. *p<0.1 **p<0.05 ***p<0.01

Table A.14: Effect of a Facebook Feed Equating the Congruence Scale of Online News Consumed

	Browsing Congruence Scale, Std. Dev.	Affective Polarization, Std. Dev.	Feeling Thermometer, Degrees
	(1)	(2)	(3)
FB Congruence Scale, Std. Dev.	0.331*** (0.070)	0.113** (0.057)	2.172 (1.551)
Control Group Diff. in Congruence: Other Sources - FB	-0.18		
Effect Required to Equate Congruence	-0.544		
Effect of Equating Congruence		-0.062	-1.112
Observations	1,072	1,071	1,030

The table shows how affective polarization would have changed if news consumed through Facebook had the same congruence scale as other news consumed. The calculation is based on the following steps. Column (1) shows the effect of the congruence scale of news exposed to in Facebook on the congruence scale of news sites visited. The congruence scale measures the mean slant of all news an individual was exposed to, multiplied by (-1) for liberal participants. Columns (2) and (3) show the effect of the congruence scale of news exposed to in Facebook on the affective polarization index and the feeling thermometer outcome. All the regressions are IV regression with the treatment as the instrument and controlling for the covariates specified in section 3.6. *Control Group Diff. in Congruence* shows that in the control group there is a difference of 0.18 standard deviations in the congruence scale between news sites visited through Facebook and other news sites visited. *Effect Required to Equate Congruence* shows that the congruence scale of the Facebook feed has to increase by 0.54 standard deviations to equate the congruence scale of sites visited through Facebook and other sites visited. It is calculated by dividing the second row by the first row. *Effect of Equating Congruence* shows how affective polarization and the feeling thermometer would have decreased if the congruence scale in the feed would have increased by 0.54 and sites visited through Facebook had the same congruence scale as other sites visited. Data includes all posts participants in the extension sub-sample were exposed to between the baseline and endline surveys. Robust standard errors. *p<0.1 **p<0.05 ***p<0.01

Table A.15: Estimations Decomposing the Segregation in News Exposure

	Subscribed OLS (1)	Exposed IV (2)	Usage OLS (3)
Offered	1.457*** (0.064)		
Offered * Pro	0.464*** (0.096)		
Subscribed		0.010*** (0.001)	
Subscribed * Pro		0.006*** (0.002)	
Counter-Att. Treatment			-19.174* (10.989)
Pro-Att. Treatment			0.477 (11.659)
Outlet and Ind. FE	X	X	
Baseline Controls			X
Observations	11,984	11,984	1,498

This table displays the regressions used to decompose the gap in exposure to pro-attitudinal posts in the pro-attitudinal treatment and exposure to counter-attitudinal posts in the counter-attitudinal treatment. In column (1), all potential outlets (the four pro-attitudinal and four-counter-attitudinal outlets defined for each participant) and participants are pooled. The dependent variable is the number of outlets participants subscribed to and the independent variable is the full interaction of whether the outlet was offered in the experiment and whether the outlet is pro-attitudinal. In column (2), all potential outlets and participants are pooled in an IV regression. The dependent variable is the share of posts from the outlet the participant was exposed to among all posts in the participant's Facebook feed and the independent variable is the full interaction of whether the participant subscribed to the outlet and whether the outlet is pro-attitudinal. Subscription to an outlet is instrumented with the outlet being offered in the experiment. In column (3), the dependent variable is the number of posts observed by the participant and the independent variable is whether the participant was assigned to the pro-attitudinal or counter-attitudinal treatment. The regression controls for Facebook visits before the intervention. In the first two columns, where observations are at the individual by outlet level, standard errors are clustered at the individual level and in the third column robust standard errors are used. *p<0.1 **p<0.05 ***p<0.01

Table A.16: Association between Consumers' Ideology and the Slant of Posts in the Facebook Feed

	Slant		Mean Slant	Slant		Mean Slant
	(1)	(2)	(3)	(4)	(5)	(6)
Conservative Ideology	0.363*** (0.019)	0.140*** (0.009)	0.124*** (0.005)	0.096 (0.063)	-0.008 (0.008)	-0.027 (0.018)
Level of Observation	URL	URL	Ind*Outlet	URL	URL	Ind*Outlet
Data = All Domains	X	X	X			
Data = Potential Outlets				X	X	X
Outlet FE		X	X		X	X
Observations	226,999	226,999	61,124	20,341	20,341	1,435

This table shows the correlation between ideology and the news individuals were exposed to in their Facebook Feed. In columns (1), (2), (4), (5), each observation is a link to an article which appeared in a participant's Facebook feed and the dependent variable is the slant of articles consumed. In columns (3) and (6), the data is aggregated at the participant by outlet level and the dependent variable is the mean slant of articles consumed. The slant of each article is calculated according to the DW-Nominate score of all congress members who shared the article on Facebook or Twitter. Each URL is only counted once for each congress member who shared it. The URLs are matched with Facebook URLs by first finding the redirected URL, and then trimming the protocol from the URL (e.g. https://) and any query within most URLs. To increase power, instead of focusing only on two weeks, the data spans from the intervention until the end of April 2018. Columns (1)-(3) include all links displayed in the Facebook feed. These columns show that there is a clear correlation between ideology and the links observed. Column (2) shows that this correlation holds even with domain fixed effects. This is not surprising as a conservative is likely to have more conservative friends, who are likely to share more conservative articles within an outlet. Columns (4)-(6) focus only on posts from the eight potential outlets which could be offered to each participant and control for the list of potential outlets. Column (4) shows that even within the outlets included in the experiment there is a correlation between ideology and the slant of articles. This could be explained by both the tendency of individuals to subscribe to outlets that fit their ideology, and by Facebook's tendency to display more posts from outlets that match one's ideology. Column (5) controls for outlet fixed effects and shows that there is no correlation between consumer's ideology and the article slant within an outlet among the potential outlets defined in the experiment. Column (6) shows that the result is robust to aggregating the data at the individual by outlet level and leads to the same conclusion. Standard errors are clustered at the individual level. * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

Table A.17: Primary Outcomes Using Different Index Methods

(a) Political Opinions					
	(1)	(2)	(3)	(4)	(5)
Liberal Treatment	−0.004 (0.005)	−0.005 (0.017)	−0.038 (0.083)	−0.004 (0.007)	−0.003 (0.005)
Conservative Treatment	−0.001 (0.005)	0.022 (0.017)	−0.058 (0.084)	0.006 (0.007)	0.004 (0.005)
Cons. - Lib. Treatment	0.003 (0.005)	0.027 (0.017)	0.010 (0.007)	−0.021 (0.052)	0.006 (0.005)
Index Method	Standrd	Inv-Cov	Inv-Cov, All Obs	Inv-Cov	Inv-Cov, All Obs
Negative Weights	-	Include	Include	Exclude	Exclude
Observations	17,128	9,251	17,128	9,251	17,128

(b) Affective Polarization			
	(1)	(2)	(3)
Pro-Att. Treatment	0.008 (0.012)	0.008 (0.014)	0.002 (0.010)
Counter-Att. Treatment	−0.024** (0.012)	−0.031** (0.014)	−0.022** (0.010)
Pro-Att. Treat. - Counter-Att. Treatment	0.032*** (0.012)	0.038*** (0.014)	0.024** (0.010)
Index Method	Standard	Inv-Cov	Inv-Cov All Obs
Observations	16,894	13,616	16,894

This table estimates the effect of the treatment on the political opinions index using different summary indexes. Column (1) uses equal weights for all outcomes included in the index. Column (2) uses inverse covariates weights and includes participants that have no missing observations for any of the measures. In column (3), inverse-covariance weights are used for all participants with non-missing outcomes, for participants with missing outcomes the weights are renormalized to sum to 1, such that an outcome measure is created for all participants who have at least one non-missing outcome. Columns (4) and (5) repeat columns (2) and (3) with non-negative weights replaced with zero and all weights renormalized to sum to 1. The specification and controls are described in Section 3.6.

Table A.18: Effect of the Treatment on Behavioral and Attitudinal Polarization Measures

	All	Affective	Behavior
Pro-Att. Treatment	0.007 (0.014)	0.008 (0.012)	−0.002 (0.018)
Counter-Att. Treatment	−0.027* (0.014)	−0.024** (0.012)	−0.010 (0.018)
Counter-Att. Treatment - Pro-Att. Treat.	0.033** (0.014)	0.032*** (0.012)	0.008 (0.019)
Controls			
Observations	17,159	16,894	16,640

This table estimates the effect of the treatment on the political opinions index (first table) and affective polarization index (second table). Column (1) uses equal weights for all outcomes included in the index. Column (2) uses inverse covariates weights and exclude participants that have missing observations for any of the measures. In column (3), inverse-covariance weights are used for all participants with non-missing outcomes, and the weights are renormalized to sum to 1 for participants with missing outcomes, such that an outcome measure is created for all participants who have at least one non-missing outcome. The inverse-covariance method creates negative weights when constructing the political opinions index. Columns (4) and (5) repeat columns (2) and (3) with non-negative weights replaced with zero and all weights renormalized to sum to 1. The specification and controls are described in Section 3.6. Both tables use robust standard errors. *p<0.1 **p<0.05 ***p<0.01

Table A.19: Effect of the Treatment on Self-reported Familiarity and Accurate Political Knowledge Outcomes

	Heard Michael Cohen (1)	Heard Clark Shooting (2)	Heard Louis Farrakhan (3)	Heard Clinton Speech (4)	Correct Russian Influence (5)	Correct Wall Built (6)	Correct Trump Target (7)	Correct Tax Cut (8)
Liberal Treatment	-0.004 (0.006)	0.005 (0.007)	-0.004 (0.006)	0.008 (0.008)	0.001 (0.005)	0.013 (0.009)	-0.003 (0.010)	0.002 (0.006)
Conservative Treatment	-0.002 (0.006)	0.001 (0.007)	-0.002 (0.006)	0.021*** (0.008)	0.008 (0.005)	-0.001 (0.009)	-0.009 (0.010)	0.001 (0.006)
Cons. Treat - Lib. Treat	0.00 (0.01)	-0.00 (0.01)	0.00 (0.01)	0.01 (0.01)	0.01 (0.01)	-0.01 (0.01)	-0.01 (0.01)	-0.00 (0.01)
Expected Effect	Lib Treat	Lib Treat	Cons Treat	Cons Treat	Lib Treat	Lib Treat	Cons Treat	Cons Treat
Observations	17,128	16,929	17,128	16,962	15,740	13,518	11,838	15,249

This table estimates the effect of the treatment on eight knowledge outcomes. All the outcomes are binary. *Michael Cohen* and *Louis Farrakhan* are whether the participant did not mark “Never heard of” when asked for their favorability ratings of the individuals. *Clark Shooting* is whether the participant heard that Stephon Clark was shot and killed by police officers in Sacramento. *Clinton Speech* is whether the participant heard that Hillary Clinton suggested many white women voted for Trump since they took their voting cues from their husbands. *Russian Influence* is agreement with “the Russian government tried to influence the 2016 presidential election”. *Wall Built* is disagreement with “the US has recently started building a new border wall at the US-Mexico border”. *Trump Target* is disagreement with “President Trump is a criminal target of Robert Mueller’s investigation”. *Tax Cut* is agreement with “most people will receive an income tax cut, salary increase or bonus under the new tax reform law”. All regressions control for party affiliation, ideology, vote, age, age squared, whether the participant follows the news and whether the participant stated they know the name of their representative in congress. Robust standard errors. *p<0.1 **p<0.05 ***p<0.01

Table A.20: Effect of the Treatment on Exposure to Words on the Facebook Feed

	Michael Cohen (1)	Clark Shooting (2)	Louis Farrakhan (3)	Clinton Speech (4)
Liberal Treatment	4.303*** (1.429)	1.216*** (0.367)	0.168 (0.125)	0.034 (0.043)
Conservative Treatment	1.627* (0.922)	0.161 (0.273)	0.408*** (0.108)	0.077** (0.033)
Cons. Treat - Lib. Treat	-2.68* (1.41)	-1.06*** (0.34)	0.24* (0.14)	0.04 (0.05)
Expected Effect	Lib. Treat	Lib. Treat	Cons. Treat	Cons. Treat
Observations	1,683	1,683	1,683	1,683

This table estimates the effect of the treatment on topics appearing in the participant's Facebook feed. *Michael Cohen* is the number of times the expression "Michael Cohen" appeared. *Clark Shooting* is the number of times the expression "Stephon Clark" appeared. *Louis Farrakhan* is the number of times the expression "Louis Farrakhan" appeared. *Clinton Speech* is the number of times the word Clinton appeared along with the word vote and either the word India or the word husband. All regressions control for party affiliation, ideology, vote, age, age squared, whether the participant follows the news and whether the participant stated they know the name of their representative in congress. Data is from participants who kept the extension installed for at least two weeks. The data is not limited to the first two weeks after the extension was installed since different participants installed the extension at different dates and the purpose of the table is to test whether the participants were exposed to specific events when they occurred. Robust standard errors. *p<0.1 **p<0.05 ***p<0.01