American Economic Review 2021, 111(3): 831–870
https://doi.org/10.1257/aer.20191777

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

By Ro'ee Levy*

Does the consumption of ideologically congruent news on social media exacerbate polarization? I estimate the effects of social media news exposure by conducting a large field experiment randomly offering participants subscriptions to conservative or liberal news outlets on Facebook. I collect data on the causal chain of media effects: subscriptions to outlets, exposure to news on Facebook, visits to online news sites, and sharing of posts, as well as changes in political opinions and attitudes. Four main findings emerge. First, random variation in exposure to news on social media substantially affects the slant of news sites that individuals visit. Second, exposure to counter-attitudinal news decreases negative attitudes toward the opposing political party. Third, in contrast to the effect on attitudes, I find no evidence that the political leanings of news outlets affect political opinions. Fourth, Facebook's algorithm is less likely to supply individuals with posts from counter-attitudinal outlets, conditional on individuals subscribing to them. Together, the results suggest that social media algorithms may limit exposure to counter-attitudinal news and thus increase polarization. (*JEL* C93, D72, L82)

In 2019, more than 70 percent of American adults consumed news on social media, compared to fewer than one in eight Americans in 2008.[1] Based on Pew surveys, Facebook is the dominant social media platform for news consumption, and

* Massachusetts Institute of Technology (email: roeelevy@mit.edu). Stefano DellaVigna was the coeditor for this article. I am deeply grateful to Ebonya Washington, Joseph S. Shapiro, and Dean Karlan for their guidance and support throughout this project. I thank three anonymous referees for their insightful comments. I also thank Hunt Allcot, Eran Amsalem, Ian Ball, Dirk Bergemann, Leonardo Bursztyn, Alex Coppock, Oren Danieli, Eduardo Fraga, Matthew Gentzkow, Matthew Graham, Eoin F. McGuirk, Alexey Makarin, Gregory J. Martin, Martin Mattsson, Brendan J. Nyhan, Rohini Pande, David Rand, Dominik Rehse, Oren Sarig, David Schönholzer, Katherine Wagner, Jaya Y. Wen, and Mor Zoran, along with seminar participants at Bar Ilan, Berkeley, Booth, Cornell, IDC, Haifa, Hebrew University, Microsoft Research, NBER Digitization, Northwestern, NYU, Tel Aviv, UCLA Anderson, UCSD, Yale, and the ZEW conference on the economics of IT, for their valuable comments. Financial support from the ISPS Field Experiment's Initiative, the Tobin Center for Economic Policy, the Yale Department of Economics, the Yale Program in Applied Economics and Policy, the Yale School of Management, and David Rand is greatly appreciated. All errors are my own. The experiment was determined to be exempt from human subjects review by the Yale Institutional Review Board. The experiment is registered at the AEA RCT registry (ID 0002713).

† Go to https://doi.org/10.1257/aer.20191777 to visit the article page for additional materials and author disclosure statement.

[1] The 2008 figure is based on the Pew Research Center 2008 Biennial Media Consumption Survey. The 2019 figure is based on the Pew Research Center American Trends Panel Wave 51, July 2019.

"among millennials, Facebook is far and away the most common source for news about government and politics" (Mitchell, Gottfried, and Matsa 2015, p. 8). As social media becomes a major news source, there are growing concerns that individuals are exposed to more pro-attitudinal news, defined as news matching their ideology, and as a result, polarization increases (Sunstein 2017).

In this paper, I test whether these concerns are warranted. I analyze the effects of exposure to pro- and counter-attitudinal news outlets by conducting a large online field experiment randomizing exposure to news on Facebook, and by collecting survey, browsing, and social media data.

To motivate the experiment, I first provide descriptive statistics on online news consumption. I show that news sites visited through social media, and specifically Facebook, are associated with more segregated, pro-attitudinal, and extreme news, compared to other news sites visited.

I recruited American Facebook users to the experiment using Facebook ads. After completing a baseline survey, participants were randomly assigned to a liberal treatment, a conservative treatment, or a control group. Participants in the liberal and conservative treatments were asked to subscribe to up to four liberal or conservative outlets on Facebook, respectively (e.g., *MSNBC* or *Fox News*), by clicking a "Like Page" button embedded at the end of the survey.[2] Remarkably, in each treatment, approximately one-half of the participants complied by subscribing to at least one outlet. When individuals subscribe to an outlet on Facebook, posts shared by the outlet may subsequently appear in their Facebook feed. A post usually contains the story's headline and often includes a link to the full news story on the outlet's website.

I designed the experiment to have high external validity. A nudge offering subscriptions to outlets is very common on social media and participants could have subscribed to any of these outlets, at no cost, without the intervention. Besides the offer, the experiment 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 outlets appeared in the participants' Facebook feeds. Finally, participants decided whether to skip, read, click, or share posts. As a result, the effect of the intervention is almost identical to the experience of millions of Americans who subscribe to popular news outlets on Facebook.

I estimate the effect of the intervention on exposure to news in the Facebook feed, news sites visited, news shared, political opinions, and affective polarization, defined as negative attitudes toward the opposing political party. Affective polarization is a primary outcome of interest since this measure of polarization has been increasing (Iyengar and Krupenkin 2018), and there are concerns over its implications for governance, accountability of elected officials, and even labor markets (Iyengar et al. 2019).

To measure subscriptions to outlets on Facebook and posts shared, I asked participants to log in to the survey using their Facebook account. To measure exposure to news in the Facebook feed and visits to news sites, I developed a Google Chrome

[2] To simplify terminology, throughout the paper I will describe the action of "liking" a page of a news organization as subscribing to an outlet on Facebook.

extension and asked a subset of participants who took the survey on a computer using Chrome to install it. To estimate the effect on opinions and attitudes, I invited participants to an endline survey approximately two months after the intervention. My sample is composed of 37,494 participants who completed the baseline survey. 34,592 of those participants provided access to the posts they shared for at least two weeks, 1,835 installed the extension for at least two weeks, and 17,635 took the endline survey.

This paper has four main findings. First, exposure to news on social media substantially affects online news consumption. Following increased exposure to posts from the randomly offered outlets, participants visited the websites of the outlets, even when the outlets did not match their ideology. Visiting the websites had a substantial effect on the mean slant of participants' overall online news consumption. The difference between the intention-to-treat (ITT) effects of the liberal and conservative treatments on the slant of news sites visited in the two weeks following the intervention is 14 percent of the difference in the slant of sites visited by liberals and conservatives in the control group.

Various economic theories explain why individuals optimally prefer news that matches their ideology (Gentzkow, Shapiro, and Stone 2015). However, I find that news consumption strongly responds to an exogenous shock to the feed, meaning that individuals often consume news incidentally, and do not fully re-optimize their browsing behavior to keep the slant of the news sites they visit constant. The results imply that social media algorithms can substantially alter news consumption habits and that while social media is associated with pro-attitudinal news, individuals are willing to engage with counter-attitudinal news when it is made more accessible on social media.

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 index includes questions such as how participants feel toward their own party and the opposing party, i.e., a "feeling thermometer." When estimating the effects on polarization, I redefine the treatments as pro- and counter-attitudinal. For example, a counter-attitudinal treatment is a liberal treatment assigned to a conservative participant or a conservative treatment assigned to a liberal participant. The ITT and treatment-on-treated (TOT) effects of the counter-attitudinal treatment on the affective polarization index, compared to the pro-attitudinal treatment, are -0.03 and -0.06 standard deviations, respectively. The TOT effect should be interpreted as the effect on individuals who subscribe to new outlets when nudged to subscribe. Comparing each treatment to the control group suggests that the effect on polarization is driven by the counter-attitudinal treatment but this result should be interpreted cautiously since participants in the control group were more likely to complete the endline survey (there is no differential attrition between the two treatment arms).

I compare the results to existing benchmarks by focusing on the feeling thermometer questions. The experiment's ITT and TOT effects decreased the difference between participants' feelings toward their own party and the opposing party by 0.58 and 0.96 degrees on a 0–100 scale over two months, respectively. For comparison, based on the American National Election Survey (ANES), this measure of affective polarization increased by 3.83–10.52 degrees between 1996 and 2016.[3]

Third, in contrast to the effect on attitudes, I do not find evidence that the slant of news outlets affects political opinions. The effect of the liberal and conservative treatments on a political opinions index focusing on issues and political figures covered during the study period is small in magnitude, precisely estimated, and not statistically significant.

The paper's fourth finding is that Facebook's algorithm may limit exposure to counter-attitudinal news. I show that participants in the counter-attitudinal treatment were exposed to substantially fewer posts from the outlets they subscribed to in the intervention, compared to participants in the pro-attitudinal treatment.

Combined, the results paint a complicated picture. On the one hand, Facebook's algorithm seems to filter counter-attitudinal news, probably since it attempts to personalize news based on the user's behavior and perceived interests. While it is not possible to estimate the effect of specific posts filtered by the algorithm, I show that exposure to counter-attitudinal news decreases affective polarization. This suggests that social media algorithms may be increasing polarization. On the other hand, this paper also shows that individuals are willing to engage with counter-attitudinal news, and social media platforms provide a setting where a subtle nudge can substantially diversify news consumption and consequently decrease affective polarization.

This paper contributes to the literature on social media and news consumption. In his seminal book *The Filter Bubble*, Eli Pariser warned that the "era of personalization is here" (Pariser 2011, p. 19). However, recent reviews concluded that "we lack convincing evidence of algorithmic filter bubbles in politics" (Guess et al. 2018, p. 12). Papers in this literature typically estimate segregation in online news based on cross-sectional analysis of browsing behavior (Gentzkow and Shapiro 2011; Flaxman, Goel, and Rao 2016; Peterson, Sharad, and Iyengar 2019; Guess forthcoming). Since they lack social media data, these papers cannot measure segregation *within* one's social media feed. One exception is a paper analyzing Facebook data, arguing that exposure to counter-attitudinal news shared by *friends* is mostly limited by individual choices and not by algorithmic ranking (Bakshy, Messing, and Adamic 2015). The paper analyzes large data but does not exploit exogenous variation. I advance the literature by generating experimental variation in subscriptions to outlets and collecting data on exposure to posts from those outlets. This allows me to decompose the mechanisms limiting exposure to counter-attitudinal news and demonstrate the existence of a filter bubble, i.e., that Facebook's algorithm is more likely to expose individuals to news matching their ideology, conditional on subscription.

My findings contribute to the literature on social media and polarization by generating variation in the main mechanism through which social media is suspected to increase polarization: the distance between individuals' ideology and the slant of their news consumption. Related papers show that the internet and Facebook

[3] The increase in polarization depends on the weights and the respondents included in the sample. When using the ANES face-to-face sample for consistency (Boxell, Gentzkow, and Shapiro 2018), the increase is 3.83 degrees. When including also the 2016 web sample (Iyengar et al. 2019), the increase is 10.52 degrees. The ANES top-codes the thermometer at 97 degrees. The results are almost exactly the same when I top-code the results.

may increase polarization (Lelkes, Sood, and Iyengar 2015; Allcott et al. 2020), but based on demographics, they may not be the primary driver in the rise of polarization (Boxell, Gentzkow, and Shapiro 2018).[4] These papers focus on the reduced-form effect of social media and do not identify the causal effect of pro- or counter-attitudinal news. Indeed, a recent review argues that "it is far from clear … that partisan news actually causes affective polarization" (Iyengar et al. 2019, p. 135). To the best of my knowledge, this paper provides the first experimental evidence that counter-attitudinal news decreases affective polarization and thus demonstrates that nudges diversifying social media news exposure can be effective.

This study also contributes to a well-established literature on media persuasion by randomly assigning subscriptions to news outlets. Survey experiments (e.g., Coppock, Ekins, and Kirby 2018) and papers with quasi-experimental designs (e.g., DellaVigna and Kaplan 2007) find that individuals are persuaded by the news they consume.[5] While in many contexts field experiments are considered the gold standard for estimating causal effects, field experiments estimating the effects of media outlets are not common. One notable exception is a study randomizing subscriptions to the *Washington Post* and *Washington Times*, which does not find an effect on opinions but is limited by a relatively small sample size (Gerber, Karlan, and Bergan 2009). This paper studies a different setting, social media, and shows how the unique features of this setting affect news exposure. Focusing on social media also allows me to analyze engagement with news and quantify the effect of news exposure.

Methodologically, this paper contributes to a growing literature conducting online media-related experiments (Bail et al. 2018, Chen and Yang 2019, Allcott et al. 2020, Jo 2020, Mosquera et al. 2020) by demonstrating how an experiment can exploit social media's existing infrastructure to gradually distribute news to participants in a natural setting. In contrast to most online experiments, participants were not asked to consume any content or continue complying with the treatment over time, not did they receive frequent reminders of the experiment. The natural, unobtrusive intervention means that it is unlikely that experimenter effects drive the study's result. To precisely detect the small effects that are expected as a result of a subtle intervention, I collect a sample size that is an order of magnitude larger than most other related experiments.

I. Background: Facebook

This study focuses on Facebook since it is the dominant social media platform, used by seven out of ten American adults. Most of these users visit Facebook several times a day,[6] and the platform accounts for 45 percent of all time spent on social media (Williamson 2018). Despite its prominence, Facebook has been understudied, especially compared to Twitter (Guess et al. 2018).

[4] Other studies estimating the effect of social media on political behavior include Bursztyn et al. (2020); Müller and Schwarz (2020); and Enikolopov, Makarin, and Petrova (2020). See Zhuravskaya, Petrova, and Enikolopov (2020) for a recent review.

[5] Other studies estimating media effects on political outcomes include Chiang and Knight (2011); Gentzkow, Shapiro, and Sinkinson (2011); Durante, Pinotti, and Tesei (2019); and Okuyama (2020). See Strömberg (2015) for a review.

[6] Facebook usage is based on the Pew Research Center January 2019 Core Trends Survey.

The most important Facebook feature is the news feed, where users scroll through a list of posts curated by Facebook's algorithm. Posts in the feed are typically shared by the user's Facebook friends, shared by Facebook pages the user subscribes to (likes), or are sponsored (advertisements shared by pages to promote content). The posts may include text, video, pictures, and links.

Facebook is a very popular source for news consumption. In 2019, 52 percent of Americans reported getting at least some of their news on Facebook, more than the share of Americans getting news on all other social media platforms combined.[7] While this study focuses on the United States, understanding Facebook's influence has global implications. A Reuters Institute survey found that in 37 out of 38 middle- and high-income countries surveyed, more than 20 percent of the population consumed news through Facebook weekly (Reuters Institute 2019). A paper analyzing the survey's data concluded that Facebook "reaches the widest international audience of any media organization in our sample" (Kennedy and Prat 2019, p. 10).

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 content on social media.

II. Design and Data

This section summarizes the experimental design, data, and empirical strategy. The design of the experiment is also presented in Figure 1.

A. *Experimental Design*

I recruited American adults to the experiment in February to March 2018 using Facebook ads. 978,628 people saw the ads, 87,648 people clicked the links in the ads, and approximately one-half of those began the survey. For more details on the ads, see online Appendix Section A.1.1. Individuals who clicked the ads were directed to the survey landing page, where they reviewed the consent form and began the survey by logging in using their Facebook account.

After logging in, and before treatment assignment, four *potential* liberal outlets and four *potential* conservative outlets were defined for each participant. The same eight potential outlets were defined for each participant unless a participant already subscribed to one of the outlets in baseline, in which case it was replaced with an alternative outlet, to ensure only new outlets would be offered. Toward the end of the survey, participants were randomly assigned to a liberal treatment, a conservative treatment, or a control group, with the randomization blocked by participants' self-reported baseline ideology.[8] Participants in the conservative treatment were

[7] Calculation based on the Pew Research Center American Trends Panel Wave 51.

[8] Respondents were asked where they position themselves on a 7-point ideological scale, with an additional option of "I haven't thought about it much." Each block is composed of three sequential participants who chose the same answer. The first participant in a block was randomly assigned to one of the three treatment groups, the second

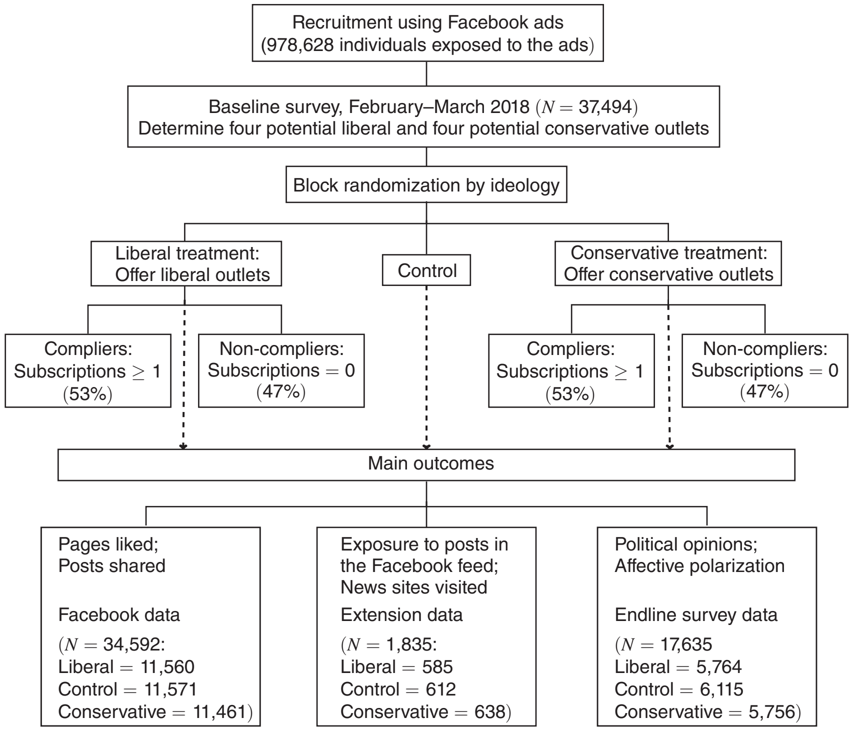

FIGURE 1. EXPERIMENTAL DESIGN

offered to subscribe (like) to their four potential conservative outlets and participants in the liberal treatment were offered to subscribe to their four potential liberal outlets. Participants in the control group were not offered any outlets.

I nudged participants to subscribe to the outlets by explaining that subscribing could expose them to new perspectives. Participants were not required to subscribe to any outlet and did not receive monetary compensation for subscribing. The intervention did not provide exclusive access to these outlets, and any individual can subscribe to these outlets on Facebook, regardless of the intervention. Since participants were logged into their Facebook account when taking the survey, the offer to subscribe was integrated within the survey, and the only action required by participants was to click the standard Like Page button. Facebook users often encounter this button, for example when Facebook suggests pages they may be interested in or when outlets promote their page. Online Appendix Figure A.1 provides an example of the intervention.

After participants subscribed to an outlet by "liking" its Facebook page, posts from the outlet appeared in their feeds, among many other posts, according to Facebook's algorithm. Participants decided 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. Due to the simple common intervention, the organic nature of any subsequent effect, and the fact that participants were not reminded of the

participant was randomly assigned to one of the two remaining groups, and the third participant was assigned to the remaining group.

intervention, experimenter effects are unlikely to play a large role in explaining the effects, at least compared to similar studies.[9] Because individuals can subscribe to outlets on Facebook at no cost and no monetary incentives were provided, the intervention is scalable.

B. *The Setting: Media Outlets and the News Environment*

The primary liberal outlets offered in the experiment were *HuffPost, MSNBC, The New York Times*, and *Slate*. The primary conservative offered outlets were *Fox News, The National Review, The Wall Street Journal*, and *The Washington Times*. The news outlets were chosen to ensure participants are offered a diverse set of popular outlets (*Fox News* and the *New York Times* are two of the three most popular news pages on Facebook) with a clear ideological slant. Online Appendix Table A.1 displays the full list of the primary and alternative outlets offered.

Figure 2 shows that the men and women mentioned most often in posts shared by the eight primary outlets and the two main alternative outlets are political figures. Unsurprisingly, President Trump is the dominant figure mentioned. Political stories that made headlines during the study period can be observed in the figure: Trump's alleged affair with Stormy Daniels, Robert Mueller's investigation, and the negotiation with North Korea's leader, Kim Jong-un. The figure also demonstrates that liberal outlets focused on scandals related to the presidency and mentioned Michael Cohen, Stormy Daniels, Scott Pruitt, and Vladimir Putin more often than conservative outlets.

C. *Data Collection and Subsamples*

Experiment Data.—The analysis of the experiment relies on three datasets: self-reported survey data, Facebook data, and browser data. This is among the first studies combining experimental variation with social media and news-related browsing data. Table 1 presents the main datasets and subsamples analyzed.

Survey Data: The endline survey measures self-reported political opinions, affective polarization, and changes in self-reported news consumption habits. A total of 17,635 participants took the endline survey and constitute the *endline survey subsample*.

Facebook Data on Pages Liked and Posts Shared: Participants logged in to the survey using their Facebook account, through a Facebook app created for the project. They were asked to provide separate permissions to access the pages they subscribe to and posts they share. Providing permissions was voluntary, they could be revoked at any time, and were revoked automatically approximately two months after participants logged in to a survey. I observe all posts shared or pages liked until permissions are revoked. Since baseline subscriptions were required to define the

[9] Participants were asked at the end of the survey what they think is the purpose of the study. Online Appendix Section C.1 shows that participants understood the study was about media and politics and that there do not appear to be dramatic differences between the answers of participants in the pro- and counter-attitudinal treatments.

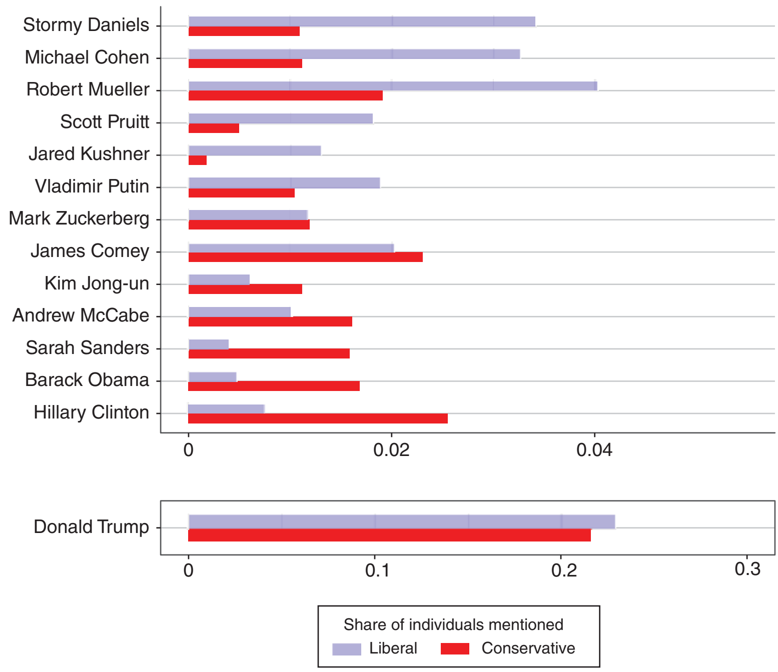

Figure 2. Figures Mentioned in Posts Shared by Outlets during the Study Period

Notes: This figure shows the men and women mentioned most often in posts shared by the eight primary outlets and two main alternative outlets between February 28 and April 25, 2018, the median dates the baseline survey and endline survey were taken. Approximately 32 percent of posts with text mentioned a name. The x-axis is the share of times an individual was mentioned in a post by one of the liberal outlets (top bars) and by one of the conservative outlets (bottom bars), of all mentions of individuals. To fit all the figures on the same scale, the x-axis is broken for Donald Trump, who is by far the most dominant person mentioned. The figures were identified using the Spacy Natural Language Processing algorithm and post-processing names (e.g., removing possessives). Names that appear in only one outlet are excluded. If only a last name is mentioned, it is associated with the dominant first and last name combination when such a combination exists. To simplify the graph, the names "Trump" and "Donald Trump" are determined to be the same individual, even though "Trump" could refer to other members of President Trump's family.

potential outlets, participants who did not provide initial permissions to access their subscriptions are excluded from the baseline sample.[10]

Data on posts shared are used to estimate the effect of the intervention on political behavior. I exclude posts sharing photos, albums, music, and events. The remaining posts typically include text with a link or an embedded video. Since posts shared are observable to the participant's social network or the general public, sharing posts can have a direct cost to the reputation of the participant. Approximately 92 percent of baseline participants provided access to the posts they shared for at least two full weeks following the intervention constituting the *access posts subsample*.

[10] While providing permissions 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 revoked permissions after the intervention are included in the baseline sample.

Table 1—Samples, Data Sources, and Outcomes

Sample	Data sources	Number of participants and retention	Main outcomes
Baseline sample	Baseline survey; Facebook data on participants' subscriptions to outlets	37,494 (all participants)	Subscriptions to outlets in the intervention (compliance)
Access posts subsample	Facebook data for participants who provided permissions to access their posts and subscriptions for at least two weeks	34,592 (94 percent of participants who provided permissions in baseline)	Subscription to outlets over time; posts shared by participants
Extension subsample	Browser data for participants who installed the extension for at least two weeks	1,835 (81 percent of participants who installed the extension in baseline)	Exposure to posts in the Facebook feed; news sites visited
Endline survey subsample	Endline survey, approximately two months after baseline	17,635 (47 percent of participants who completed the baseline survey)	Political opinions; affective polarization

Notes: This table describes the main sample and subsamples analyzed along with the data sources, the number of participants, and the main outcomes. The subsamples and data are described in Section IIC. The outcomes are described in Section IID.

Extension Data on Browser Behavior and the Facebook Feed: 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 intervention. A total of 2,262 of the 8,084 participants who were offered the extension installed it. I focus on 1,835 participants who kept the extension installed for at least two weeks and constitute the *extension subsample*.

The Facebook feed data are used to analyze news exposure by estimating how often participants were exposed to posts from outlets on Facebook. I observe the posts that participants saw when they used their computer mouse to scroll their feed. I do not observe whether a post is a sponsored advertisement, but identify suspected ads as posts in the feed from pages participants did not subscribe to and posts appearing in the feed repeatedly. I attribute a post to a news outlet if it was created by the outlet's Facebook page or contains a link to the outlet's domain.[11] While the variation generated by the experiment is in subscriptions to the outlets' Facebook pages, my analysis includes news articles shared by the participants' friends, to accurately capture total exposure to news outlets on Facebook.

The browsing behavior data is used to estimate the effect on the news sites participants visited. The extension can greatly reduce measurement error, compared to self-reported estimates of news consumption, as individuals' self-reported media habits may be more polarized than their actual news consumption (Guess, Nyhan, and Reifler 2017).

[11] To match URLs with news outlets, I first convert over ten million URLs to their final endpoint, following redirects. This is required since many links on Facebook are based on URL-shortening services such as tinyurl.com.

The extension data were only collected when participants used a computer while being signed into their Chrome account. In practice, individuals often use Facebook and browse news sites on a mobile device or at work, where they may use a different browser. Therefore, the estimates for the number of posts participants were exposed to in their feed and the number of sites they visited are lower bounds.[12]

Additional details on the survey, Facebook, and extension data can be found in online Appendix Sections A.1, A.2, and A.3, respectively.

Subsamples: The datasets define three separate subsamples. To maximize power, when analyzing the effects on opinions and attitudes, I focus on the *endline survey subsample* and when analyzing media outcomes, I focus on the *extension subsample* and the *access posts subsample* (or their overlap). Online Appendix Table A.2 presents descriptive statistics on the subsamples and shows that the extension subsample is more liberal and older, as would be expected when excluding participants who took the survey on a smartphone. The share of compliers is greater in the extension subsample, which assists in detecting treatment effects despite the smaller sample size.

External Data.—

Outlets: I measure the slant of news at the outlet level, the common method used in the literature. I determine an outlet's slant according to a dataset by Bakshy, Messing, and Adamic (2015) defining the slant of 500 news domains based on the self-reported ideology of Facebook users sharing articles from the domains. Using this definition, a completely liberal outlet has a slant of approximately –1, a middle-of-the-road outlet has a slant of approximately 0, and a completely conservative outlet has a slant of approximately 1. I use this measure of slant since it is based on 2014 data, and thus more recent than other common measures, and since it covers a large number of online news outlets. The dataset correlates well with other measures of slant (e.g., Gentzkow and Shapiro 2010). I refer to outlets in this dataset as *leading news outlets*. I exclude from the dataset several popular domains which are clearly not news outlets or that serve mostly as portals, and merge several outlets that are associated with multiple domains. I manually determine the Facebook pages of leading outlets by searching for pages with names similar to each outlet's domain. Facebook pages were found for 370 outlets.

Comscore Browsing Data: To provide descriptive statistics on news consumption outside the experimental sample, I analyze the 2017 and 2018 Comscore WRDS Web Behavior Database Panel (Comscore 2018). Each observation in the dataset is a domain visited by a specific computer along with the referral domain. I merge this dataset with the list of leading news outlets. The Comscore data provide several advantages. The combined 2017 and 2018 samples include 94,342 individuals

[12] 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 participants in the extension subsample who provided a numerical answer under 1,000, approximately 72 percent of news links were clicked on a computer, so it is likely that most, but not all data are collected for these participants.

who visited at least one news site. Previous studies have shown that the panel is representative of online buyers in the United States (De los Santos, Hortaçsu, and Wildenbeest 2012). Finally, the data have been collected for previous years and used by other researchers (Gentzkow and Shapiro 2011), allowing me to estimate changes in news consumption over time. I classify the channels through which visitors reached websites as social, search, or direct visits. Facebook is by far the dominant referral source in the social category.

For more details on the outlet and Comscore datasets, see online Appendix Sections A.4 and A.5, respectively.

D. *Outcomes*

Media.—I measure subscriptions to outlets on Facebook, exposure to news in the Facebook feed, news sites visited, and posts shared, using the following outcome measures. First, I estimate the direct effect of the experiment according to the number of times participants engaged with the *potential outlets* (the four liberal outlets and the four conservative outlets defined for each participant). For example, I measure the number of posts that participants observed from their potential liberal and conservative outlets in their feed. Second, I measure the mean slant of all *leading news outlets* with which participants engaged. 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 consumed, multiplied by (-1) for liberal participants. This scale has a higher value when individuals consume more extreme news matching their ideology. Fourth, I measure the *share of counter-attitudinal news*, defined as the share of news from counter-attitudinal outlets among all news from pro- and counter-attitudinal outlets.[13]

Opinions and Attitudes.—I analyze the effects of news exposure on two primary outcomes: political opinions and affective polarization. For both outcomes, an index is composed by taking an average of all the valid non-missing index components and then standardized by subtracting the control group mean and dividing by the control group's standard deviation.

The political opinions index is composed of 20 survey questions focusing on domestic political issues and political 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. Each outcome variable is defined such that a higher value is associated with a more conservative opinion and then standardized.

The affective polarization index is composed of five outcomes. First, I use the feeling thermometer questions (*feeling thermometer*). Second, participants were 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 were asked a similar question on the following statement: "I think it is important to consider the perspective of

[13] Counter-attitudinal outlets are defined as outlets in the two most liberal quintiles for conservatives or the two most conservative quintiles for liberals. For more details on the definitions, see online Appendix Section B.1.

Democrats/Republicans" (*consider perspective*).[14] Fourth, participants were asked if they think the Democratic 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 attitudes toward the party associated with the participant's ideological leaning and attitudes toward the opposing party, a typical measure of affective polarization. Fifth, participants are asked if they would feel very upset (2), somewhat upset (1), or not upset at all (0) if they had a son or daughter who married someone from the opposing party (*marry opposing party*).[15] Each outcome variable is defined such that a higher value is associated with more polarization and then standardized.

E. *Empirical Strategy*

When estimating the effects of the intervention on engagement with the liberal and conservative outlets, the slant of news with which participants engaged, and their political opinions, I compare the liberal and conservative treatments. When measuring the effects on polarization and engagement with pro- and counter-attitudinal outlets, 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- and counter-attitudinal treatments. This strategy follows the study's pre-analysis plan, discussed in online Appendix Section B.2.

Liberal and Conservative Treatments: I estimate the following ITT regression:

$$Y_i = \beta_1 T_i^L + \beta_2 T_i^C + \alpha X_i + \varepsilon_i, \tag{1}$$

where $T_i^L, T_i^C \in \{0,1\}$ is whether participant i is assigned to the liberal or conservative treatment, respectively. 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). To increase power, I control for the following set of covariates, X: self-reported ideology, party affiliation, approval of President Trump, ideological leaning, age, age squared, gender, and baseline questions measuring political opinions that are similar to questions used in the endline survey. Online Appendix Section B.3 describes the control variables. When estimating the effect on media outcomes, I only control for the outcomes in

[14] Both statements are based on an empathy index developed by Robb Willer, Jamil Zaki, and Emily Reit, loosely based on the Interpersonal Reactivity Index (Davis 1980).

[15] Participants stating in the endline survey that they are Republicans or Democrats were asked how they would feel if they had a son or daughter who married a Democrat or Republican, respectively. Participants who did not identify with either party were asked about one of the parties randomly. I asked participants only about the opposing party since I was concerned they would find it odd to state how upset they would be if they had a son or daughter who married someone from their own party. However, conditioning the question on an endline variable could potentially bias the result. For example, if some participants were affected by the counter-attitudinal treatment, and as a result, no longer identified with their party, they were less likely to be asked about the opposing party in endline and the average participant asked about the opposing party would be slightly less moderate in this treatment arm. I include this measure in the affective polarization index since it is the only social-distance measure in the index, it is included in the pre-analysis plan, and any bias is expected to go against the direction of my findings. Online Appendix Table A.13 shows that the results are robust to excluding this measure from the index.

baseline, when they exist. All regressions use robust standard errors unless noted otherwise.

Pro-Attitudinal and Counter-Attitudinal Treatments: I estimate the following ITT regression:

$$Y_i = \beta_1 T_i^A + \beta_2 T_i^P + \alpha X_i + \varepsilon_i \,, \tag{2}$$

where $T^A \in \{0, 1\}$ is whether the participant was assigned to the counter-attitudinal treatment, defined as a liberal treatment assigned to a conservative participant or a conservative treatment assigned to a liberal participant. Similarly, $T^P \in \{0, 1\}$ is whether the participant was assigned to the pro-attitudinal treatment, defined as a liberal treatment assigned to a liberal participant or a conservative treatment assigned to a conservative participant. The term X is 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. Here, $\beta_1 < \beta_2$ tests whether individuals become more polarized when assigned to pro-attitudinal news, compared to counter-attitudinal news.

I determine whether participants are liberal or conservative (their ideological leaning) according to the party they identify with or lean toward. If participants do not identify with either the Democratic or Republican Party, their ideological leaning is defined according to their self reported ideology, and if they do not identify as liberal or conservative, it is defined according to the candidate they supported in the 2016 elections.[16]

F. *Balance and Attrition*

Table 2 presents descriptive statistics for participants in the baseline sample and shows that the sample is balanced. Online Appendix Table A.3 presents a balance table according to whether the treatment matched the participant's ideology (pro- or counter-attitudinal), and shows that the sample is balanced along the redefined treatment arms as well. The sample size in this table is slightly smaller because it excludes participants for whom an ideological leaning cannot be defined.

Similar to other opt-in panels, the sample is not nationally representative. Participants tend to be more liberal than the US population and, as expected, more participants say that they get most of their news on social media (18 percent), compared to the national population (13 percent). The share of female participants and the average age is similar to the US population. Self-reported exposure to news on Facebook in line with one's views is similar to US Facebook users. Overall, the sample seems at least as representative as samples of Mechanical Turk users (Berinsky, Huber, and Lenz 2012).[17]

[16] Approximately 3 percent of participants do not identify with the Republican or Democratic Party, do not self-identify as liberals or conservatives, and did not support Trump or Clinton. They are excluded from the analysis when analyzing the effect of the pro- and counter-attitudinal treatments. The effect on affective polarization is robust to including only participants who identify with or lean toward the Democratic or Republican Party.

[17] One advantage of the sample is that Facebook users are not experienced, semi-professional survey takers. Participants were asked in the endline survey how many additional surveys they completed in the past month, the

Table 2—Balance Table, Liberal, and Conservative Treatments

	Mean			Difference		
Variable	Sample $N = 37{,}494$	United States	Facebook users	Control - Lib.	Control - Cons.	Cons. - Lib.
Baseline survey						
Ideology (−3, 3)	−0.61	0.17		0.01	0.01	0.00
Democrat	0.38	0.35	0.30	0.01	0.00	0.01
Republican	0.17	0.28	0.21	−0.01	0.00	−0.01
Independent	0.37	0.32	0.35	−0.00	−0.00	−0.00
Vote support Clinton	0.53			−0.00	−0.00	−0.00
Vote support Trump	0.26			0.00	−0.00	0.01
Feeling therm., Rep.	29.07	43.06		0.11	0.25	−0.13
Feeling therm., Dem.	46.99	48.70		0.40	0.46	−0.06
Difficult pers., Rep. (1, 5)	3.13			0.02	0.00	0.02
Difficult pers., Dem. (1, 5)	2.39			−0.00	0.01	−0.01
Facebook echo chamber	1.18		1.12	−0.00	−0.00	0.00
Follows news	3.35	2.42		0.01	0.01	−0.00
Most news social media	0.18	0.13		−0.00	0.00	−0.00
Device						
Took survey mobile	0.67			−0.01	−0.00	−0.01
Facebook						
Female	0.52	0.52	0.55	−0.01	−0.00	−0.00
Age	47.69	47.30	42.86	0.22	−0.13	0.35
Total subscriptions	474			5.15	9.04	−3.89
News outlets slant (−1, 1)	−0.18			0.00	0.00	0.00
Access posts, pre-treat.	0.98			0.00	0.01	−0.00
Attrition						
Took followup survey	0.47			0.03	0.03	−0.00
Access posts, 2 weeks	0.92			0.00	0.01	−0.01
Extension install, 2 weeks	0.05			0.00	−0.00	0.00
F-test				1.20	0.89	1.05
P-value				[0.21]	[0.64]	[0.39]

Notes: This table presents descriptive statistics, along with the difference between participants assigned to each treatment arm. *Vote support* is the share of participants who voted for or preferred the candidate. *Difficult pers.* is whether participants find it difficult to see things from Democrats'/Republicans' point of view. *Facebook echo chamber* is whether the opinions participants see about government and politics on Facebook are in line with their views always or nearly all the time (3), most of the time (2), some of the time (1), or not too often (0). *Follows news* is whether participants follow government and politics always (4), most of the time (3), about one-half of the time (2), some of the time (1), or never (0). *Total subscriptions* is the number of Facebook pages participants subscribed to in baseline. *News outlets slant* is the slant of news outlets subscriptions. *F*-tests are calculated by regressing the treatment on the pre-treatment variables, with missing values replaced with a constant and an indicator for a missing value. Data sources for the United States and Facebook population are specified in online Appendix Section C.4.1.

Table 2 and online Appendix Table A.3 also test for differential attrition among the three subsamples. The access posts and extension subsamples have low attrition rates compared to baseline take-up (as shown in Table 1) and very small differences in attrition by treatment arm. Therefore, their results are unlikely to be affected by attrition.[18] However, more participants completed the endline survey in the control

median answer is 1 and the mean answer is 7. For comparison, a 2014 study found that the median Mechanical Turk worker reported participating in 20 academic studies in the *week* before the question was asked (Rand et al. 2014).

[18] 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

group (49 percent), compared to the liberal (46 percent) and conservative (46 percent) treatment arms. The differential attrition mostly stems from a small share of participants in the conservative and liberal treatments who did not complete the final screen of the baseline survey after they encountered the intervention.

Online Appendix Tables A.4 and A.5 present balance tables for the endline survey subsample and show that despite the attrition, the two treatment arms and control group are similar on observables. Participants in the pro-attitudinal treatment who completed the endline survey are *not* substantially more polarized in baseline than participants in the counter-attitudinal treatment. Moreover, there is no differential attrition between the conservative and liberal treatments and no differential attrition between the pro- and counter-attitudinal treatments. When estimating the effect on the primary endline survey outcomes, I compare the two treatment arms to each other to mitigate concerns over differential attrition. Still, it is possible that attrition could affect the results.

G. *Compliance*

Throughout the analysis, I focus on ITT estimates. To measure the effect of complying with the treatment, defined as subscribing to at least one offered outlet, I also analyze TOT estimators by regressing the dependent variable on compliance with each treatment and instrumenting compliance with the random treatment assignment. Since the intervention only offers new outlets to participants, defiers do not exist in this experiment.[19] Because compliance is defined as liking an outlet when it was offered, always-takers do not exist either.[20] If compliers are more likely to engage with the outlets and be affected by them, perhaps because they are more interested in the content, the TOT is expected to be larger than the average treatment effect.

In the entire baseline sample, 59 percent of participants who were offered pro-attitudinal outlets complied with the pro-attitudinal treatment and subscribed to at least one outlet, compared to 48 percent of participants offered counter-attitudinal outlets. Table 3 shows that participants were more likely to subscribe to outlets they are familiar with, to outlets with a perceived ideology similar to their own ideology, and to outlets they perceive as more moderate. Online Appendix Table A.6 presents descriptive statistics on the compliers by treatment arm and shows that liberals, women, and participants who subscribe to more outlets on Facebook were more

intervention. However, this minimal difference seems to be random since a small statistically significant difference between the groups in providing permissions already existed before the intervention.

[19] Defying the experiment would mean unsubscribing from an offered outlet, but participants are only offered outlets to which they are not already subscribed. There are rare cases where I only observe a partial list of outlets in baseline and as a result, a participant could have been offered an outlet she already subscribed to and "unliked" the outlet's page instead of "liking" it. However, I estimate that I observed a partial list of outlets for less than 1 percent of participants and I do not have evidence that participants unsubscribed from outlets as a result of the intervention.

[20] In a handful of cases, participants subscribed to potential outlets even when the outlets were not offered, possibly since the survey included questions about these outlets. However, these cases are extremely rare and therefore, I am not defining them as compliance for simplicity. When focusing on the two weeks following the intervention instead of immediate compliance, an always-taker would be defined as a participant who would subscribe to a potential outlet in that period, regardless of the intervention. In the control group, only 0.2 percent and 0.5 percent of participants subscribed to a potential conservative or liberal outlet, respectively, in the two weeks following the intervention.

Table 3—Compliance with the Treatments

	(1)	(2)
Conservative treatment, conservative ideology	0.513	
	(0.008)	
Liberal treatment, conservative ideology	0.349	
	(0.008)	
Conservative treatment, liberal ideology	0.541	
	(0.006)	
Liberal treatment, liberal ideology	0.623	
	(0.006)	
Know slant		0.230
		(0.006)
Outlet ideology, absolute value (standard deviation)		−0.047
		(0.003)
Ideological distance (standard deviation)		−0.083
		(0.002)
Controls	X	X
Observation unit	Ind.	Ind. × outlet offered
Observations	36,728	97,937

Notes: This table estimates the association between participants' characteristics and compliance with each treatment arm. In column 1, the dependent variable is whether the participant subscribed to at least one offered outlet and the independent variable is the interaction of participant's ideological leaning and her treatment assignment. The reference group is the control group where there are no compliers. In column 2, the data is pooled such that each observation is a participant and an outlet offered. The dependent variable is whether the participant subscribed to the outlet. The independent variables are based on the outlet's perceived ideology according to the participant, where ideology is measured on a 7-point scale from extremely liberal to extremely conservative with an additional option of "do not know." *Ideological distance* is the standardized difference between the participant's self-reported ideology and the outlet's perceived ideology. Both regressions control for age, age squared, gender, and the set of potential outlets defined for a participant, and column 2 also controls for outlet fixed effects. Column 1 use robust standard errors and column 2 clusters standard errors at the individual level.

likely to comply with both treatments. To test whether participants open to new ideas comply more often with the treatments, I use two questions from a brief measure of the big five personality traits (Gosling, Rentfrow, and Swann 2003), along with self-reported certainty in political opinions, and exposure to counter-attitudinal news in baseline. Based on these measures, compliers with the counter-attitudinal treatment are slightly more open than non-compliers, but the differences are not large (0.12–0.19 standard deviations) and some of these differences exist to a lesser degree when comparing compliers and non-compliers among participants assigned to the pro-attitudinal treatment.

This section deals with immediate compliance with the intervention, which is especially useful when interpreting the TOT effects. However, the experiment is designed to allow participants to opt-out of news content at any stage. They could always unsubscribe from the offered outlets or ignore posts from the outlets appearing in their feed. Therefore, the effects found will probably be driven by participants who decide to consume the content offered when it becomes accessible. This feature increases the external validity of the results because these participants are often the policy-relevant population, as they are more likely to engage with the offered outlets in other circumstances as well.

III. Descriptive Analysis: Segregation in Online News Consumption

Is the rise of social media associated with a change in news consumption patterns? In this section, I present descriptive statistics on segregation in social media and online news. I calculate two main measures: isolation and segregation.

Isolation measures whether conservatives and liberals visit different websites. It is defined as the difference between exposure to conservatives in websites visited by conservatives and exposure to conservatives in websites visited by liberals. Exposure to conservatives is the share of conservatives visiting each set of websites. Intuitively, if conservatives tend to visit websites visited by many other conservatives, while liberals tend to visit websites visited by few conservatives, the isolation measure is higher. To make the measure comparable to estimates by Gentzkow and Shapiro (2011), I aggregate visits at the daily level and use the adjusted leave-out estimator of isolation.

Segregation is defined as the scaled standard deviation of the slant of news sites visited by participants. To keep this measure in the unit interval, the slant of outlets is normalized to range from 0 to 1. A higher value means that there is a greater difference in the slant of news consumed by two random individuals. The measures are formally defined in online Appendix Section B.1.

A. *Segregation in Online News*

I find that news consumed through social media is more segregated and extreme than news consumed through other channels. Figure 3 shows that in the 2017–2018 Comscore sample, the segregation index is 0.18 for news sites visited through search engines, 0.21 for news sites visited directly, and 0.28 for news sites visited through social media. I cannot precisely calculate the isolation measure for the Comscore panel since individual ideology is not observed. Instead, in panel B of Figure 4, I analyze isolation based on the extension subsample and show that isolation is greater in news sites visited through Facebook, compared to news sites visited through other means. The analysis is based on participants assigned to the control group and includes data from the first eight weeks after the extension was installed. The full results are presented in online Appendix Table A.7, which also shows that total segregation is similar in the extension and Comscore datasets.

The increased segregation for news sites visited through social media could stem from the composition of the individuals using social media to consume news. Online Appendix Table A.8a presents the segregation among the 8,882 individuals in the Comscore sample who visited multiple news sites through Facebook and through other sources. As all the individuals in this group consume news through both channels, the comparison better isolates the effect of the medium. While the share of news sites visited through Facebook is much greater among these individuals (26 percent), sites visited through Facebook remain substantially more segregated than sites visited through other means.

Panel A of Figure 5 presents the distribution of the mean slant of news consumption for these individuals and shows that news sites visited through Facebook are more extreme. When visiting news sites through Facebook, 57 percent of individuals consume news that is on average more conservative than the *Wall Street Journal*

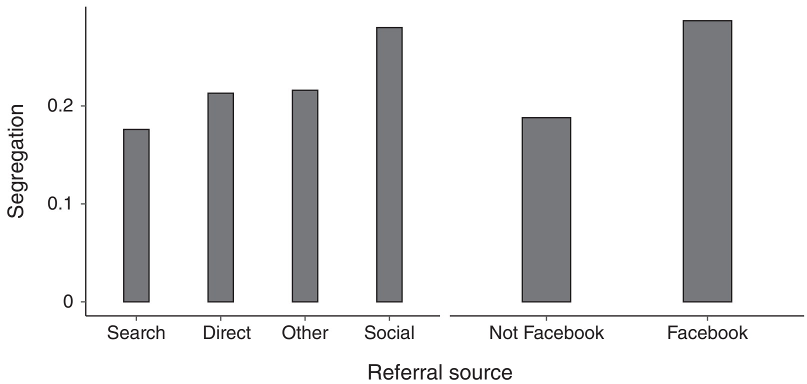

Figure 3. Segregation in News Sites Visited by Referral Source, Comscore Data

Notes: This figure displays the segregation in news sites visited by referral source. The definition of the segregation measure is discussed in Section III. Online Appendix Section A.5 defines the websites composing each channel.

Source: 2017–2018 Comscore data

or more liberal than the *Washington Post*, and when visiting news sites through other sources, 39 percent of individuals consume such partisan news.[21]

Panel B of Figure 5 shows a clear correlation between the consumers' ideology and the slant of their news consumption. I proxy for ideology using the share of donations to Republican candidates in the consumers' zip codes in the 2016 and 2018 election cycles, based on FEC data. The slope for news consumed through Facebook is steeper than the slope for news consumed through other sources, indicating that sites visited through Facebook tend to better match the consumers' ideology.

Has segregation in online news consumption increased? In the extension sample, the *segregation* index for all online news is 0.20 when I define the slant of outlets based on Bakshy, Messing, and Adamic (2015) and 0.23 when I define the slant based on the ideological leaning of participants (Peterson, Sharad, and Iyengar 2019). These segregation levels are similar to a value of 0.25 found by Peterson, Sharad, and Iyengar (2019) using 2016 data from the Wakoopa toolbar and substantially larger than a value of 0.11 found by Flaxman, Sharad, and Rao (2016) using 2013 Bing toolbar data. To compare the *isolation* index to previous estimates, I use visit-level measures of isolation (row 6 in online Appendix Table A.9b), which give more weight to individuals who visit more news sites. The isolation of browsing behavior in the extension sample is 0.22, similar to a value of 0.21–0.24 calculated by Peterson, Sharad, and Iyengar (2019) and larger than a value of 0.07–0.08 calculated by Gentzkow and Shapiro (2011). One limitation with this comparison is that while I attempt to make the samples comparable, each study still analyzes the data slightly differently. In online Appendix Table A.8b, I provide a cleaner comparison

[21] *Washington Post* and the *Wall Street Journal* are in the thirty-sixth and sixty-third percentile of the Bakshy, Messing, and Adamic (2015) dataset. When using the twenty-fifth and seventy-fifth percentile, which are similar to the *Boston Globe* and *Fox News*, 19 percent of individuals consume news that is on average more extreme than those outlets when visiting news sites through Facebook and 5 percent consume such extreme news when visiting news sites through other sources.

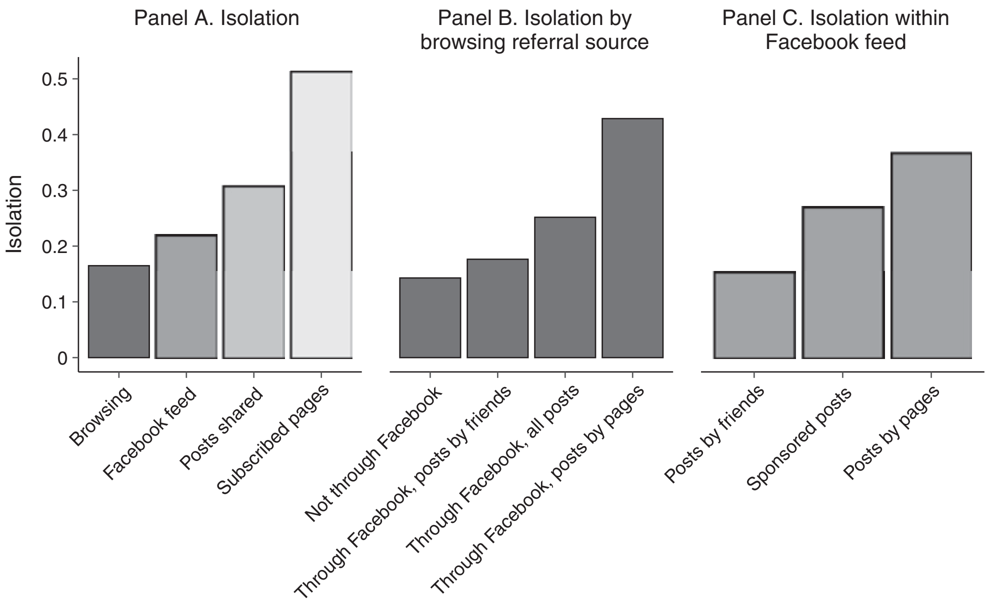

Figure 4. Isolation by Medium, Extension Data

Notes: This figure displays the isolation of news participants engaged with. A higher value means liberals and conservatives were more likely to engage with different news outlets. Panel A shows the isolation measure for news sites participants visited, posts that appeared in their feed, posts they shared, and news outlets they subscribed to on Facebook. Panel B compares isolation values for news sites visited through different sources. Panel C compares different types of posts in the Facebook feed. The figure analyzes data from control group participants in the first eight weeks after the extension was installed.

of segregation levels by comparing 2007–2008 and 2017–2018 Comscore data and do not find substantial changes in segregation.

The analysis does not lead to conclusive results regarding changes in news consumption. Segregation online may have increased, but it probably did not change dramatically. How does this result coincide with increased segregation on social media? While I find that Facebook is more segregated than other online content, and while Facebook is typically the first or second most important source for online traffic, social media still accounts for a limited share of visits to news sites. For an average individual in the Comscore sample, 4 percent of news sites were visited through Facebook and in the extension subsample, which only includes Facebook users, the figure is 15 percent.[22] Therefore, social media can be substantially more segregated than news consumed through other sources without dramatically changing overall segregation in online news consumption.

[22] These estimates may underestimate Facebook usage since they rely on browsing activity on computers, while Facebook is more popular on mobile. For comparison, Parse.ly (2018) tracks pages viewed in thousands of sites and estimates that 16 percent of traffic related to Donald Trump in April to May 2018 was from social media and that Facebook is the largest external referral source for traffic in the law, government, and politics category.

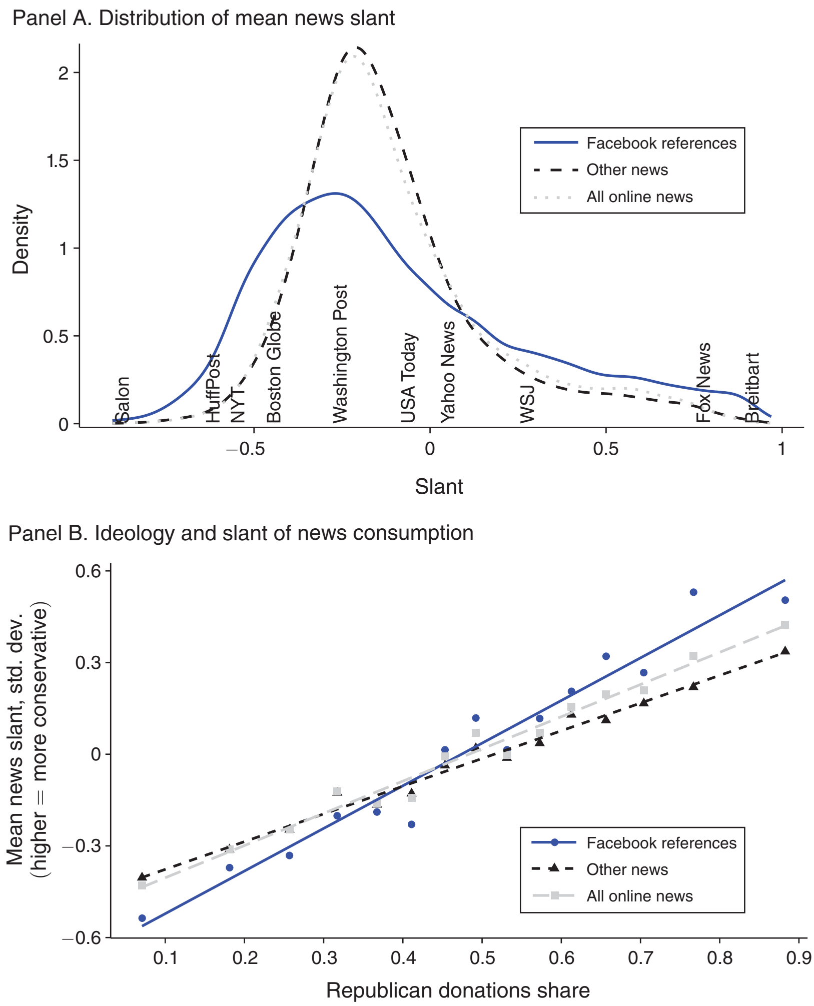

Figure 5. News Consumption in the Comscore Panel

Notes: Panel A presents the distribution of the mean slant of news sites visited (smoothing bandwidth = 0.05). Major news outlets are added to the *x*-axis for reference. The slant of each domain is based on Bakshy, Messing, and Adamic (2015). A visit is referred from Facebook if the referring domain is "facebook.com." Panel B presents a binned scatter plot. The *x*-axis is the share of Republican donations in a zip code based on FEC donation data for the 2016 and 2018 election cycles and the *y*-axis is the mean slant of news sites visited. The sample for both figures includes individuals in the 2017 and 2018 Comscore Web Behavior Database Panel who visited news sites multiple times through Facebook and through other sources.

B. *Segregation within Facebook*

Why does news consumed through Facebook tend to be more extreme and segregated? Two mechanisms that could increase segregation are homophily in social networks (an "echo chamber" effect where one's friends tend to recommend like-minded news sources) and the abundance of accessible, free media options allowing consumers to personalize their news feed. Panels B and C of Figure 4 show that the increased segregation is mostly associated with Facebook pages (the outlets participants subscribe to on Facebook) and not with Facebook friends (the social network). The isolation index is 0.14 when participants visit news sites not through Facebook, 0.18 when they visit sites through posts shared by Facebook friends, and 0.43 when they visit sites through posts shared by Facebook pages. Online Appendix Table A.7b shows that the results hold for the segregation measure as well.[23]

This descriptive analysis cannot completely isolate each mechanism, nor rule out additional mechanisms. For example, posts by both Facebook friends and Facebook pages are also affected by Facebook's algorithm, which is discussed in more detail in Section VI. Still, the analysis suggests that in order to understand segregation in social media, it is important to study the forces determining which pages appear in the Facebook feed and the effect of posts from these pages. Furthermore, posts shared by pages should not be ignored since approximately one-half of visits to news sites through Facebook in the extension subsample are through links shared by pages (row 10 in online Appendix Table A.9b).

To conclude, in a 2019 survey, 83 percent of Americans stated that one-sided news is a very big or moderately big problem on social media.[24] This section provides evidence that this concern is warranted, as it shows that news accessed through Facebook is indeed more segregated and extreme than other online news. The next section estimates the causal effects of exposure to more and less segregated news using the random variation generated by the experiment.

IV. Findings: Demand for News on Social Media

A. *Individuals Are Willing to Engage with Counter-Attitudinal News*

Figure 6 displays the effects of the pro- and counter-attitudinal treatments on engagement with the potential pro- and counter-attitudinal outlets, respectively. To keep the results comparable across media outcomes, the figure is calculated for the participants who both installed the browser extension and provided permissions to access their posts for at least two weeks. Each row in the figure is estimated by regressing engagement with the four potential pro- or counter-attitudinal outlets in the two weeks following the intervention on the pro- or counter-attitudinal treatment. The control group is the reference group. Throughout the analysis, I use linear regressions for ease of interpretation. Online Appendix Table A.10 shows that the effects on the feed, browsing behavior, and posts shared are qualitatively similar when running Poisson regressions.

The first panel of Figure 6 shows that the counter-attitudinal treatment increased the number of subscriptions to counter-attitudinal outlets by 1.42, compared to the control group. The effect is significant as the entire confidence interval is greater than zero. The increase is similar to the number of outlets participants immediately subscribed to in the intervention (1.51, not shown in the figure) since few participants unsubscribed from these outlets within two weeks.

[23] Figure 4 also provides a comparison of the isolation index in outlets individuals subscribe to on Facebook, posts they see in their feed, news sites they visit, and posts they share (panel A), and shows that isolation is highest among subscriptions.

[24] Pew Research Center American Trends Panel Wave 51, July 2019.

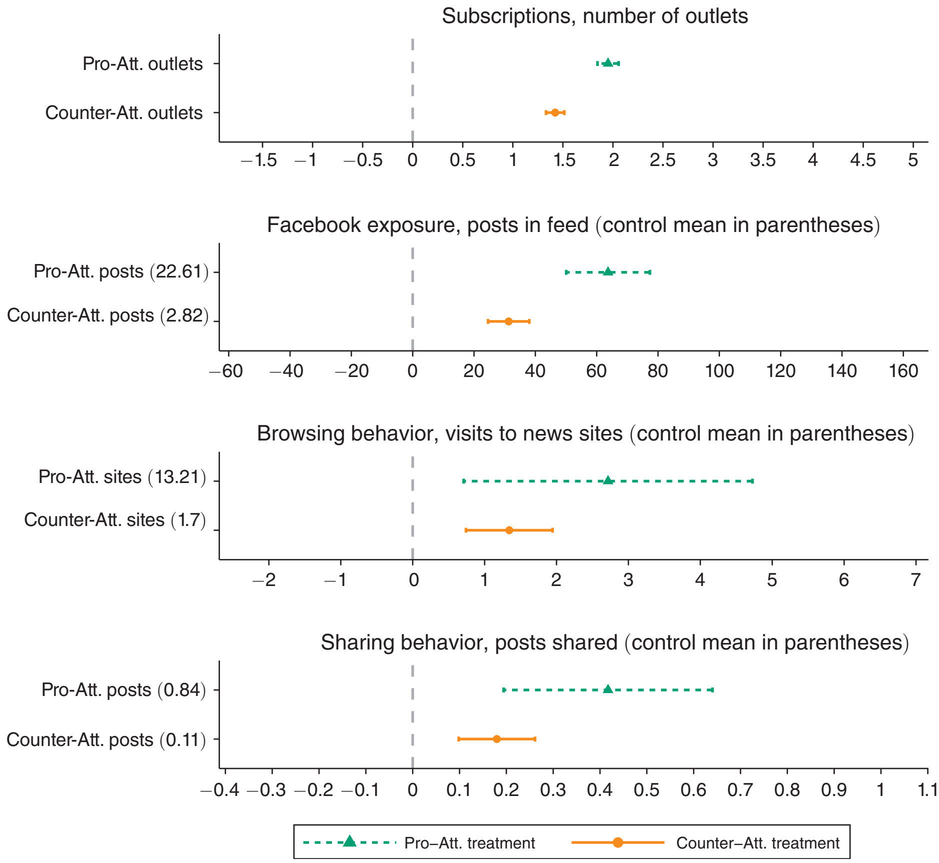

Figure 6. Effects of the Pro- and Counter-Attitudinal Treatments on Subscriptions, News Exposure, News Sites Visited, and Sharing Behavior, Two Weeks Following the Intervention

Notes: This figure shows the effect of the treatments on engagement with the offered outlets in the two weeks following the intervention. The dependent variable is engagement with either the four potential pro-attitudinal outlets or the four potential counter-attitudinal outlets and the independent variable is the treatment. Each panel presents the effect of a separate outcome. For example, in the third panel, the triangle point and dashed line present the point estimate and the confidence interval of the effect of the pro-attitudinal treatment on visits to the websites of the potential pro-attitudinal outlets, compared to the control group. The regressions control for the outcome measure in baseline if it exists. The sample includes 1,648 participants with a liberal or conservative ideological leaning who installed the extension and provided permissions to access their posts for at least two weeks. Error bars reflect 90 percent confidence intervals.

Exposure to Posts in the Facebook Feed: The second panel of Figure 6 shows that in the two weeks following the intervention, participants in the pro- and counter-attitudinal treatments were exposed to 64 and 31 additional posts from the potential pro- and counter-attitudinal outlets, respectively. For comparison, control group participants were exposed to 266 posts from leading news outlets, and 2,335 posts in total, suggesting that the intervention affected news exposure but did not take over the participants' feeds.

Online Appendix Figure A.2 shows that the effect on exposure is driven mostly by organic posts published by pages and not by sponsored posts or posts shared by friends, meaning that participants were exposed to the content directly, without commentary from their social network. To test whether participants noticed the posts, they were asked in the endline survey how often they saw news from various outlets in their Facebook feed in the past week. Online Appendix Figures A.3 and A.4 show that participants reported seeing more news from the outlets they were offered and that participants in the counter-attitudinal treatment were less likely to say that opinions they see in their feed are aligned with their views. This implies that the effect on the feed was noticeable for at least two months, and confirms that the treatment affected the large subsample of participants who completed the endline survey and not only participants who installed the extension.

News Sites Visited: The third panel of Figure 6 shows that the counter-attitudinal treatment increased total visits to the websites of the counter-attitudinal outlets by 79 percent, an ITT effect of 1.34 visits over a baseline of 1.70 visits in the two weeks following the intervention. The pro-attitudinal treatment increased the number of visits to the websites of pro-attitudinal outlets by 21 percent, an ITT effect of 2.72 visits over a baseline of 13.21.

Online Appendix Figure A.5 separately estimates the effects of the intervention on the number of visits to the outlets' websites through a link appearing in the Facebook feed and on visits not directly associated with Facebook. While there is a strong and significant effect on visits through Facebook, there also seems to be an effect on other visits, albeit the latter result is not precisely estimated. It is possible that once participants read an article on the outlets' websites, they followed links to other articles as well. Alternatively, when participants became more familiar with the new outlets, they may have started visiting those outlets even without a Facebook referral. Online Appendix Figure A.6 shows that participants were more likely to click posts appearing higher in the feed. This could occur both because participants are more curious when they just start scrolling their feed and because Facebook's algorithm ranks posts according to expected interest. Interestingly, conditional on the order of posts, participants were as likely to visit a link from an outlet they subscribed to as a result of the intervention, compared to other news outlets.

Sharing Behavior: The fourth panel of Figure 6 shows that participants not only consumed news from counter-attitudinal outlets when they appeared in their feeds, they also shared the posts. To increase power and verify that the effect is not limited to participants who installed the extension in online Appendix Figure A.7, I analyze this effect using the entire access posts subsample and show that both treatments had a significant effect on the number of posts shared by these participants. The fact that participants chose to share the posts suggests that they considered the posts important, and implies that participants expanded the treatments to their social network.

Complementing previous studies focusing on Twitter (Halberstam and Knight 2016), participants were much more likely to share pro-attitudinal posts. However, the relative effect on sharing counter-attitudinal posts compared to the control group (an increase of 105 percent) is stronger than the relative effect of the pro-attitudinal treatment (53 percent). Participants may have shared posts while commenting negatively on their content. The second panel of online Appendix Figure A.7 focuses on posts that were shared with no commentary by the participants and shows that even among these posts, the counter-attitudinal treatment had a significant effect on the number of posts shared.

B. *The Social Media Feed Strongly Affects Online News Consumption*

The previous section demonstrated that individuals engage with the potential outlets when they appear in their feed, suggesting that news is often consumed incidentally when it becomes more accessible. This raises the question of whether individuals adjust the rest of their news consumption such that the slant of their 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 consequently consume less news from the *Boston Globe*, which offers a similar perspective. To test whether the treatment affected the mean slant of all news with which participants are engaged, I focus on the conservative and liberal treatments since there are clear predictions on how these treatments would affect the slant.

Exposure to Posts on Facebook: The first panel of Figure 7 shows that when participants were randomly offered liberal or conservative outlets, their feed became substantially more liberal or conservative, respectively. The combined ITT effect of the liberal and conservative treatments equals 36 percent of the gap between the slant of the feed of liberals and conservatives in the control group, where slant is measured based on the leading news outlets dataset (participants who did not visit any news sites are excluded). The corresponding TOT effect is 47 percent. The change in slant provides a strong first stage, which is useful when analyzing the effect on political beliefs. It also allows me 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 participants re-optimize the sites they visit following an exogenous shock to their feed.

News Sites Visited: I find that individuals do *not* fully re-optimize their news consumption to keep the slant of the news sites they visit constant. The second panel of Figure 7 shows that the treatments had a strong and significant effect on the slant of news sites visited by the participants. The combined effects of the liberal and conservative treatments equal 14–19 percent (ITT-TOT) of the difference in the slant of news sites visited by conservatives and liberals in the control group. Based on the Comscore panel, the TOT effect of the liberal treatment would shift the online news diet of an individual in Pennsylvania, a swing state, to a diet similar to an individual in New York, a blue state, and the TOT effect of the conservative treatment would shift the individual's news consumption to a news diet similar to an individual in South Carolina, a red state.[25] Online Appendix Table A.11 shows that the effect on slant is robust across various subsamples (e.g., when excluding participants who did not complete the endline survey).

By combining the exposure and browsing data, I find that when the compliers' news feed became one standard deviation more conservative, the slant of the news sites they visit became 0.31 standard deviations more conservative, and the slant of the subset of sites visited through Facebook became 0.71 standard deviations more

[25] I determined the mean news consumption of each individual in Comscore's 2017 and 2018 panels based on visits to leading news outlets. Individuals who visited only one news site are excluded. The slant is then calculated at the state level for all panel members in the state.

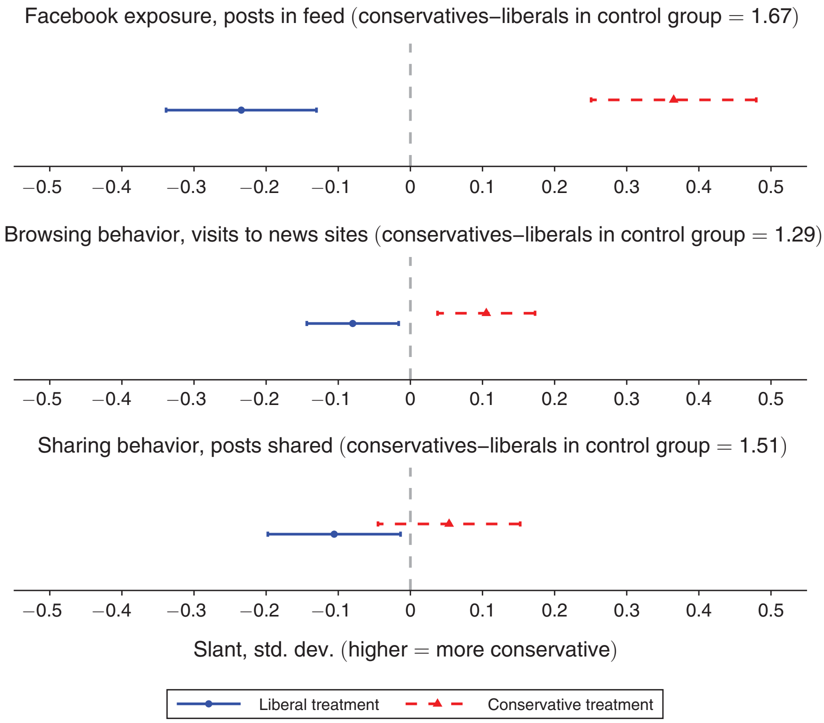

Figure 7. Effect of the Treatments on News Slant

Notes: This figure shows the effect of the liberal and conservative treatments on the mean slant, in standard deviations, of all news with which individuals engaged. In each panel, the dependent variable is the mean slant of outlets and the independent variable is the treatment. The regressions control for the outcome in baseline, if it exists. The sample includes 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.

conservative (both effects are significant at the 1 percent level). These estimates are calculated by instrumenting the slant of the posts observed in the Facebook feed with the treatment assignment. The regressions rely on the exclusion restriction that the treatments only affected the slant of sites visited through the slant of the Facebook feed. While the intervention is only expected to have an effect through the Facebook feed, the treatments could affect the feed in many ways. I am condensing the feed, a complicated object, to a scalar, the mean slant of news to which an individual was exposed. This scalar is strongly affected by the treatment assignment and has intuitive economic meaning, but other changes in the feed, not captured in this measure, could affect the news sites visited. Since these calculations rely on stronger assumptions than the ITT and TOT estimates, they should be interpreted cautiously.

To test for spillovers across news outlets, I calculate the effect of the treatments on the mean slant of all leading outlets excluding the eight potential outlets defined for each individual. Online Appendix Figure A.8 shows that the mean slant of news consumption is not strongly affected by the treatments when the potential outlets are excluded, implying that the experiment did not have large crowd-in or crowd-out effects.

Persistence: It is possible that participants were initially curious about the new outlets they were offered but quickly stopped engaging with them. Figure 8 shows that the effect of the conservative treatment on news slant, compared to the liberal treatment, declines over the first six weeks after the intervention but mostly remains positive and significant. Online Appendix Figure A.9 repeats this analysis for the first 12 weeks after the intervention. While these results should be interpreted more cautiously since a substantial number of participants did not keep the extension installed or provide permissions to access posts over this longer time period, they suggest that the effects of the experiment declined but remained significant for at least 12 weeks.

The long-term effects also alleviate concerns that experimenter effects are driving the results in this section, as it is unlikely that participants remembered which posts appeared in their feed as a result of the intervention two months after the baseline survey, assumed that the experimenter expected them to persistently visit these websites, were constantly conscious that some of their browsing behavior could be observed, and were willing to spend time visiting news sites only to leave an impression on the experimenter. Furthermore, a survey question in the endline survey suggests that most participants did not remember which outlets they subscribed to and therefore their behavior or answers are unlikely to have been driven by experimenter effects.[26]

C. *Discussion*

This section shows that people are willing to substantially change their news consumption and engage with counter-attitudinal news on social media. Online Appendix Section C.2 analyzes the content of posts participants engaged with based on the words appearing in the posts and the article sections the posts linked to (e.g., Politics, Business, or Arts). I find that a large share of content tends to be political, even when the outlets the participants engaged with were counter-attitudinal.

How do these results coincide with the previous section, which shows that news consumed through social media tends to be pro-attitudinal? If news is consumed incidentally on social media, and the Facebook feed tends to be pro-attitudinal, individuals are more likely to visit pro-attitudinal websites through social media but they will start visiting counter-attitudinal websites when they appear in their feed. Passive news consumption can also explain why Chen and Yang (2019) find that providing access to uncensored internet does not lead to consumption of censored foreign news. As long as consumers are passive, providing access to new outlets may not be sufficient to affect news consumption because consumers will continue

[26] Participants were asked "In a previous survey, we may have asked if you are interested in 'liking' Facebook news pages. Did you like a page in the previous survey?" Only 40 percent of participants in the treatment arms stated that they remembered whether they liked a page and which pages they liked. Unfortunately, many participants did not understand this question and assumed it refers to a previous question in the endline survey. Therefore, I interpret this question as providing qualitative evidence that many participants did not remember which outlets they subscribe to and not for empirical analysis. The misunderstanding probably leads to an overestimation of the number of participants who remember which pages they liked as some respondents may have remembered the previous question in the endline survey but not the outlets offered in the baseline survey. Furthermore, even among the minority of participants who understood the question and stated that they remember which pages they liked, some did not state the correct outlets.

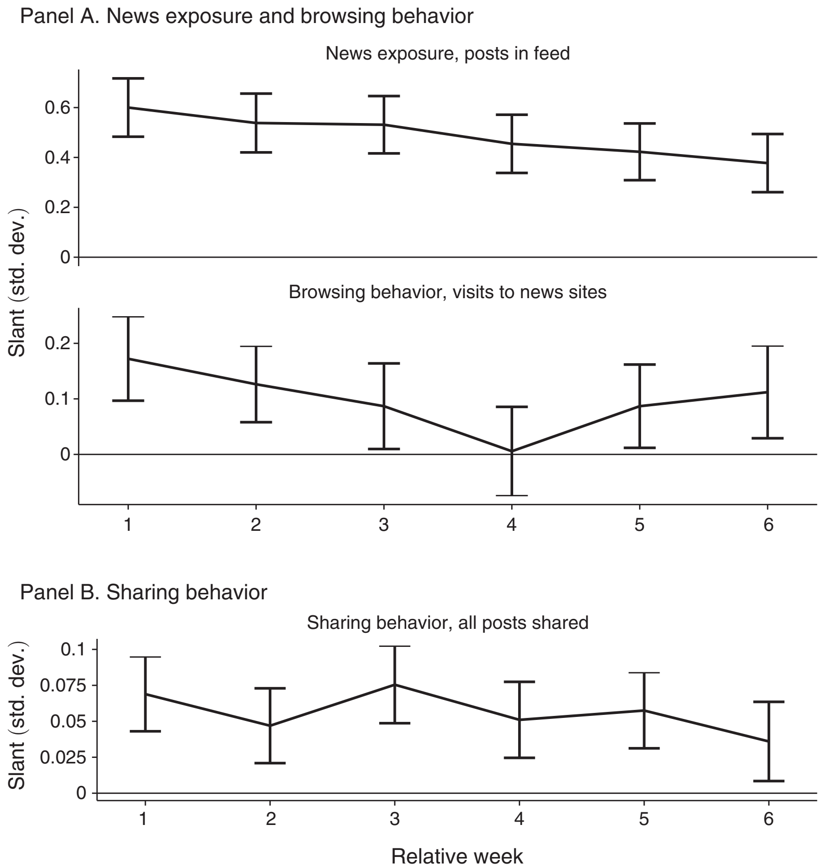

Figure 8. Effects of the Conservative Treatment on Mean Slant by Week, Compared to the Liberal Treatment

Notes: These figures show the difference between the effect of the liberal and conservative treatments on the mean slant of news engagement over time. Each panel presents a series of regressions, where the dependent variable is the slant of outlets in a specific week. The regressions control for the outcome in baseline when it exists. In the *x*-axis, relative week 1 is a full week immediately following the intervention. Panel A is based on 1,596 participants who kept the extension installed for at least six weeks following the intervention. Panel B is based on 29,131 participants who provided access to posts they shared for at least six weeks. Error bars reflect 90 percent confidence intervals.

visiting the default outlets appearing in their bookmarks, search results, or social media feeds. My intervention may have affected news consumption because it increased the salience of specific outlets and decreased the search costs required to visit them by showing them on Facebook often.

This conclusion raises concerns regarding the power of social media companies in shaping news consumption habits. The effect of the social media feed on news consumption implies that any change to the feed, due to new subscriptions or a change in the algorithm, can drastically change one's news diet. Attempts to change the feed by suggesting new content happen all the time. They can stem from companies attempting to maximize profits by increasing user engagement or originate from entities attempting to maximize political goals, such as political candidates purchasing ads or even foreign agents promoting Facebook pages to influence the American electorate.[27]

V. Findings: Opinions and Attitudes

A. *Social Media News Exposure Does Not Strongly Affect Political Opinions*

The top panel of Figure 9 shows that the treatments did not affect the political opinions index. While the point estimate has the expected sign, the effect is minimal (0.005 standard deviations), precisely estimated, and not statistically significant. The upper bound for the combined liberal and conservative treatment effects, based on a 95 percent confidence interval, is only 0.8 percent of the difference in political opinions between liberals and conservatives in the control group. Online Appendix Figure A.10 shows that the effect on each component of the political opinions index is small, and I cannot reject a null effect for any of the components.

Why did the treatments not affect political opinions even though they dramatically affected the Facebook feed of participants? In other settings, studies on persuasion found a null effect that masked substantial heterogeneity (Baysan 2020). Perhaps some participants were persuaded by the offered outlets, while for others, there was a backlash effect and opinions moved in the opposite direction of their treatment assignment. The top panel of online Appendix Figure A.11 estimates the effect of the interaction of ideology and treatment on the political opinions index and does not find evidence for a backlash effect. A second option is that the treatment did not affect political opinions since social media is still not a dominant news source, compared to television. This could explain why the results of this study differ from studies on *Fox News* (DellaVigna and Kaplan 2007, Martin and Yurukoglu 2017). Interestingly, I do not find evidence for heterogeneity based on whether participants reported getting most of their news on social media (see online Appendix Section C.3). It is also possible that the null effects are explained by the fact that the intervention lasted for two months. However, the intervention lasted long enough to affect attitudes, as discussed in the next section.

The results differ from a recent study that found a backlash effect when exposing individuals to counter-attitudinal content on Twitter (Bail et al. 2018). Differences in the experiments' design can explain the differing results. Bail et al. (2018) expose individuals to bots retweeting counter-attitudinal *views*. Individuals plausibly become more upset when exposed to opposing opinion leaders, compared to counter-attitudinal news outlets. Bail et al. (2018) also provided monetary incentives to continuously follow the bots, asked participants to disable Twitter's timeline algorithm to ensure they viewed the tweets, and included weekly surveys to verify compliance. In my setting, participants could decide whether to comply with the treatment and engage with the content. Therefore, compliers with each treatment arm are different by design and this could affect the results. Social scientists have criticized the generalizability of forced exposure media experiments since the effects

[27] For example, many ads purchased by Russian organizations in their attempt to influence the 2016 election promoted Facebook pages. Congress has published the ads and they can be found here: https://intelligence.house.gov/social-media-content/social-media-advertisements.htm.

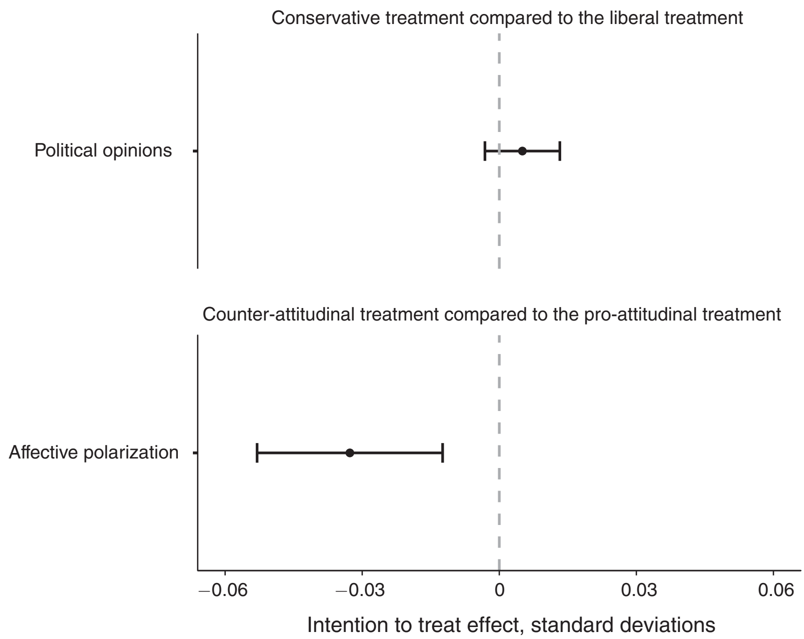

Figure 9. Effect of the Treatments on Political Opinions and Affective Polarization

Notes: This figure shows the effect of the treatments on the primary endline survey outcomes. 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 counter-attitudinal treatment on the affective polarization index, compared to the pro-attitudinal treatment. A higher value is associated with a more polarized outcome. The indices are described in Section IID and the regressions specifications are detailed in Section IIE. The panels are based on 17,635 participants who took the endline survey. Error bars reflect 90 percent confidence intervals.

found may be concentrated among individuals who would not consume the content outside the experimental setting (Hovland 1959, Bennett and Iyengar 2008). For example, conservatives who get upset when visiting msnbc.com are less likely to consume content from *MSNBC* in my setting but may consume such content when encouraged to do so by the experimenter, and this type of consumption could drive the backlash effect.

B. *Exposure to Counter-Attitudinal News Decreases Affective Polarization*

The bottom panel of Figure 9 shows that the counter-attitudinal treatment modestly decreased the affective polarization index compared to the pro-attitudinal treatment. The ITT and TOT effects are 0.03 and 0.06 standard deviations, respectively. This suggests that the concerns over more segregated news consumption are not misguided. When estimating the effect on each component of the index separately in online Appendix Figure A.12, the effect is largest on whether participants find it difficult to see things from each party's point of view.

Online Appendix Tables A.12b, A.13, and A.14b show that the result is robust to excluding covariates, dropping each of the five components of the affective polarization measures from the index one at a time, and excluding participants who already subscribed to at least one of the primary outlets before the intervention. Online Appendix Table A.15b shows that an effect is detected when focusing on the subsample of participants who completed the endline survey and installed the extension. The effect is stronger among this group, which also had higher compliance rates. Online Appendix Section C.4 shows that the effect is similar when the regressions are reweighted to match populations means in ideology, party affiliation, gender, age, and the baseline feeling thermometer measure. Online Appendix Section C.5 estimates heterogeneous effects using causal forests (Wager and Athey 2018) and shows that the predicted effect in the entire baseline sample is very similar to the effect among the endline survey subsample.

Comparing each treatment separately to the control group shows that most of the difference between the pro- and counter-attitudinal treatments stems from the counter-attitudinal treatment, perhaps because the relative effect of this treatment on engagement with the outlets was larger compared to baseline. In all specifications, the effect of the counter-attitudinal treatment is negative, statistically significant, and stronger than the effect of the pro-attitudinal treatment. However, this comparison suffers from differential attrition, due to lower attrition in the control group. Therefore, in online Appendix Table A.12, I also calculate Lee bounds for the effects of each treatment (Lee 2009). Due to the relatively small treatment effect, the bounds include a null effect. As an additional robustness test, I exclude control group participants who were recruited using the last email or ad inviting them to the endline survey (Behaghel et al. 2015). Without these participants, I compare the 45–46 percent of participants in each treatment arm who were "easiest" to recruit and attrition is similar across treatments. The results using this method are almost identical to the main specification.

I do not find evidence for substantial heterogeneity across most covariates I test for, including age, ideological leaning, baseline interest in news, and baseline exposure to counter-attitudinal news (online Appendix Section C.3). One exception is that the treatments seemed to have a stronger effect on participants who were less polarized in baseline according to the feeling thermometer question. However, this effect is significant only at the 10 percent level and more research is required on heterogeneity.

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 extension 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 a more balanced feed. All the results are based on the effect of the offered outlets over two months and could be different with longer exposure or if different outlets were offered.

The ITT and TOT effects of the counter-attitudinal treatment decrease the difference between the feeling toward the participant's party and the opposing party by 0.58 and 0.96 degrees (on a 0–100 scale), respectively. For comparison, in the past 20 years, the feeling thermometer measure increased by 3.83–10.52 degrees. An additional point of comparison is a recent experiment by Allcott et al. (2020), who found that disconnecting from Facebook for one month in the fall of 2018 decreased the feeling thermometer measure by 2.09 degrees.[28] Hence, one way to interpret these results is that almost one-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.

To estimate the effect of exposure to pro- or counter-attitudinal news on polarization, I focus on participants who installed the browser extension and completed the endline survey, i.e., the overlap between the extension and the endline subsamples. I use two summary measures for exposure to pro- and counter-attitudinal news: the share of counter-attitudinal news in the Facebook feed and the feed's congruence scale (both defined in Section IID). 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. I estimate the effect of each measure on affective polarization, and instrument the measure with the treatment assignment. Similar to the discussion in Section IVB, the IV regressions rely on the exclusion restriction that the treatment only affects affective polarization through its effect on the measure analyzed.

I find that an increase of 1 standard deviation in the share of exposure to counter-attitudinal news decreases affective polarization by 0.13 standard deviations and an increase of 1 standard deviation in the congruence scale has a similar effect. The effects are significant at the 10 percent level as the sample size is smaller when focusing on participants who both installed the extension and completed the endline survey. One challenge in studying affective polarization based on non-experimental survey data (e.g., Garrett et al. 2014) is determining whether the correlation between 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. Online Appendix Table A.16 shows that the effects of news exposure on affective polarization are approximately 26–34 percent of the coefficients obtained using a cross-sectional regression among the control group, suggesting that the correlation is both due to a causal effect and selection.

I use the effect of the Facebook feed to estimate how affective polarization would have changed if individuals were exposed to more balanced news on Facebook. I find that if the feed had an equal share of pro- and counter-attitudinal news, the difference between the feelings toward one's party and the opposing party would decrease by 3.94 degrees. For this calculation, I estimate the effect of increasing the share of exposure to counter-attitudinal news by 33 percentage points, the difference between exposure in the control group and an exposure of 50 percent. The estimation does not rely on out-of-sample predictions as the share of counter-attitudinal news was greater than 50 percent for many participants in the counter-attitudinal treatment. Using a similar exercise, I find that if the congruence of the Facebook feed equaled zero, the difference between participants' feelings toward their one party and the opposing party would decrease by 3.43 degrees.

Perhaps a balanced news feed is not a realistic counterfactual because most individuals do not consume balanced news, regardless of social media. Therefore, in

[28] Focusing on one measure decreases power. The effect I find on the feeling thermometer is statistically significant at the 10 percent level and the Allcott et al. (2020) benchmark is not statistically significant.

a second back-of-the-envelope calculation, I estimate how affective polarization would change if individuals were exposed in their Facebook feed to the same share of counter-attitudinal outlets, or the same congruence scale, as they encounter when visiting news sites not through Facebook. I find that the feeling thermometer outcome would decrease by 0.24–0.62 degrees. These calculations should be interpreted carefully since they do not take into account general equilibrium effects.[29] Nevertheless, they suggest that the Facebook feed may slightly amplify polarization.

Interpretation.—Why did the treatments affect attitudes toward political parties but not political opinions? One possibility is that participants learned new facts about the world and these facts swayed their attitudes. Based on eight pre-registered survey questions, I test whether a change in participants' knowledge could explain the effect on polarization. In online Appendix Section C.6, I do not find evidence for strong effects on knowledge.

Previous studies showed that Americans believe that members of the opposing party are more likely to hold extreme views than they actually do (Yudkin, Hawkins, and Dixon 2019), and therefore, attitudes may have changed because participants learned the opposing party is not as extreme as they thought.[30] I do not find evidence that the pro- and counter-attitudinal treatments had a significant effect on the distance between participants' baseline ideology and the perceived ideology of each party (online Appendix Figure A.4).

Another option is that exposure to pro- and counter-attitudinal news affects attitudes due to increased negative coverage (Levendusky 2013). This explanation predicts that pro-attitudinal outlets would increase negative attitudes toward the opposing party and counter-attitudinal outlets would affect consumers' attitudes toward their own party. This prediction is inconsistent with the data. I measure separately the effect of each treatment on attitudes toward each party and show in online Appendix Table A.17 that the effect of the counter-attitudinal treatment on attitudes toward the opposing party is driving the results.

An alternative explanation, consistent with the data, is that participants exposed to counter-attitudinal news learned to rationalize the opinions of the opposing party. Intuitively, participants may have learned some of the opposing party's arguments and thus understood better why that party supports certain positions. This led to more positive attitudes but did not change political opinions as long as participants did not find these arguments particularly important. In online Appendix Section D, I formalize this discussion using a model where political opinions are a weighted average of multiple beliefs and parties place different weights on beliefs.

[29] For example, it is likely that if Facebook drastically changed its feed, individuals would use other social media platforms instead. Some of this effect may be captured in the calculations since participants in the counter-attitudinal treatment used Facebook less often (as discussed in Section VI). However, with network effects, the decrease in Facebook use could be greater. The calculations also ignore the indirect effect of Facebook on news sites visited.

[30] This theory is consistent with a study by Orr and Huber (2020) who find that negative attitudes toward individuals from the opposing party decrease when information is provided about their policy position.

There could be other explanations for the change in affective polarization.[31] The literature on affective polarization is relatively new and more research is needed to pinpoint the precise mechanisms explaining how affective polarization evolves.

VI. Findings: Exposure to Pro-Attitudinal News on Social Media

The previous section shows that exposure to pro-attitudinal news affects partisan hostility, therefore it is important to understand what influences the news individuals are exposed to on social media. This section decomposes the gap in exposure to posts shared by the pro- and counter-attitudinal outlets offered in the experiment into three main forces: participants are less likely to subscribe to counter-attitudinal news outlets; Facebook's algorithm supplies fewer posts from counter-attitudinal outlets, conditional on participants subscribing to them; and participants use Facebook less often when offered counter-attitudinal outlets. The decomposition exercise is based on the following framework:

$$E_{ij} = S_{ij} A_{ij} U_i,$$

where E_{ij}, exposure, is the number of posts from outlet j that individual i was exposed to in her Facebook feed. Exposure is a product of whether individual i subscribed to outlet j (S_{ij}), the share of posts by the outlet among all posts the individual was exposed to, conditional on subscribing to the outlet (A_{ij}), and the total number of posts individual i was exposed to (U_i). I decompose the gap in exposure using the following formula:

$$\Delta E = \underbrace{S_{\Delta} A_C U_C}_{\text{Subscriptions}} + \underbrace{S_C A_{\Delta} U_C}_{\text{Algorithm}} + \underbrace{S_C A_C U_{\Delta}}_{\text{Usage}} + \underbrace{S_{\Delta} A_{\Delta} U_C + S_{\Delta} A_C U_{\Delta} + S_C A_{\Delta} U_{\Delta} + S_{\Delta} A_{\Delta} U_{\Delta}}_{\text{Combinations}}, \tag{3}$$

where for each variable, the subscript C denotes the value for the counter-attitudinal treatment and Δ denotes the difference between the pro- and counter-attitudinal treatments. *Subscriptions* is the additional counter-attitudinal posts participants assigned to the counter-attitudinal treatment would have been exposed to if they would have subscribed to the same number of outlets as participants assigned to the pro-attitudinal treatment. *Algorithm* is the additional posts these participants would have been exposed to if Facebook's algorithm would have supplied them with the same share of posts from counter-attitudinal outlets, as the share supplied when subscribing to pro-attitudinal outlets. *Usage* is the additional posts these participants would have been exposed to if they would have used Facebook as much as participants assigned to the pro-attitudinal treatment.

[31] The counter-attitudinal treatment may have mitigated tribalism, which could have decreased affective polarization (Mason 2015). Indeed, field experiments have found that strengthening partisan behavior can affect political behavior and beliefs (Gerber, Huber, and Washington 2010). I use party affiliation as a proxy for tribalism and find in online Appendix Figure A.4 that the treatments did not significantly affect this proxy. However, the point estimate of the effect on Democratic Party affiliation has the predicted sign, and I cannot reject a small effect on affiliation with the Democratic Party.

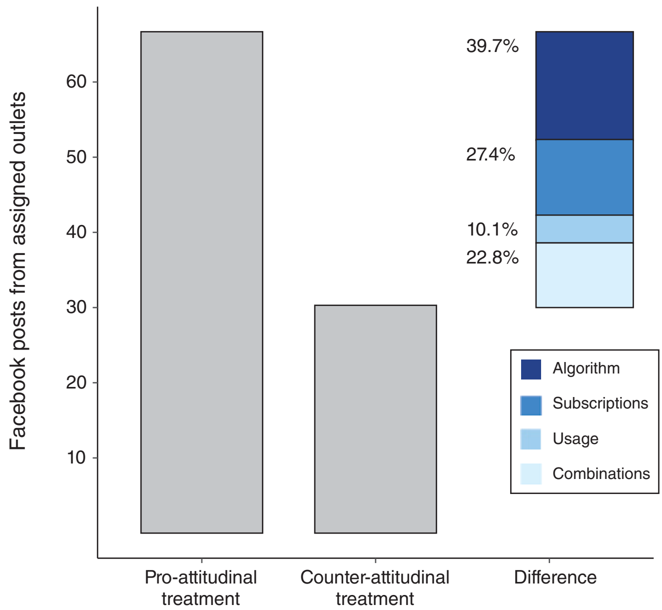

Figure 10. Decomposing the Gap between Exposure to Posts from the Offered Pro-Attitudinal and Counter-Attitudinal Outlets

Notes: This figure decomposes the gap between the number of posts participants were exposed to from the offered pro- and counter-attitudinal outlets. The *y*-axis is the number of posts seen in the feed in the two weeks following the intervention and the *x*-axis is the treatment arm. Algorithm describes the gap explained by Facebook's tendency to show participants a greater share of posts from pro-attitudinal outlets (among all posts in the feed) conditional on subscriptions. Subscriptions describes the gap explained by participants' tendency to subscribe to more offered outlets in the pro-attitudinal treatment. Usage describes the gap explained by participants' tendency to view fewer posts on Facebook (use Facebook less often) in the counter-attitudinal treatment. Combinations describe interactions between these expressions. Data are based on 1,059 participants in the pro- and counter-attitudinal treatments for which posts in the Facebook feed could be observed in the two weeks following the intervention and at least one post is observed. The calculations appear in online Appendix Section C.7.

The terms S_C and U_C are the mean number of new subscriptions and the total number of posts participants were exposed to, respectively, in the counter-attitudinal treatment. I estimate S_Δ and U_Δ by regressing the number of subscriptions and total exposure to posts on whether participants were assigned to the pro- or counter-attitudinal treatment. To estimate A_Δ and A_C, I pool the two groups of potential outlets for each participant such that each observation is a participant and either the group of pro-attitudinal outlets or the group of counter-attitudinal outlets. I then regress the share of posts that the participant was exposed to from a group of outlets (among all posts in the feed) on the full interaction of the number of new outlets the participant subscribed to and whether the group of outlets is pro-attitudinal. Since subscriptions are endogenous, they are instrumented with whether the group of outlets was randomly offered to the participant. The calculations are discussed in detail in online Appendix Section C.7 along with alternative decompositions.

Figure 10 shows that the strongest force associated with participants' increased exposure to pro-attitudinal news is the algorithm. This demonstrates that 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 (a phenomenon often described as a filter bubble). I also find evidence that participants prefer to subscribe to pro-attitudinal news outlets and that participants decrease their Facebook usage after they are offered to subscribe to counter-attitudinal outlets. The last effect is only significant at the 10 percent level and should be interpreted more cautiously. Still, it could explain why personalization is leading to segregation on social media. When consumers are exposed to more counter-attitudinal news, they may decrease their Facebook usage, and therefore, platforms have an incentive to filter counter-attitudinal news to maximize engagement. This result raises the question of whether the algorithm also personalizes content within an outlet, by showing conservatives more conservative posts shared by an outlet and liberals more liberal posts shared by the same outlet. In online Appendix Section C.7.3, I find no evidence for within-outlet personalization.

Interestingly, even though I find that the algorithm seems to be filtering counter-attitudinal posts, Section III shows that the posts control group participants are exposed to in their feed are not more pro-attitudinal than the outlets they subscribe to on Facebook. One possible explanation for the differing results is that individuals probably subscribe to outlets as a response to nonrandom nudges. If nudges typically offer pro-attitudinal outlets, then users will subscribe to these outlets often and only users who are specifically interested in opposing content will subscribe to counter-attitudinal outlets. As a result, the algorithm may filter less counter-attitudinal content.[32] The comparison to the control group descriptive statistics not only demonstrates why an experiment is necessary but also has policy implications. Adjusting the algorithm to offer more balanced news, conditional on subscription, would not make a big difference if individuals only subscribe to pro-attitudinal outlets. Therefore, to increase diversity in news exposure, nudges encouraging subscriptions to diverse outlets may also be required.

This section does *not* suggest that Facebook's algorithm intentionally increases segregation by ranking posts according to whether they match the user's beliefs, or that the interaction of the slant of an outlet and ideology of a user has a causal effect on a post's ranking. Platforms rank posts based on many signals that can be correlated with whether an outlet is counter-attitudinal, including the consumer's past engagement with the outlet, her social network, and possibly other pages she subscribes to. In other words, the effect of the algorithm also captures the behavior and perceived interests of the user. Indeed, online Appendix Section C.7.2 shows that the effect of the algorithm slightly increases over time, suggesting that engagement with content plays a role in the ranking of posts.

Personalization of news exposure is still an important departure from how news was supplied in the past. Until recently, the engagement of an individual with news, e.g., the articles she read in the newspaper or the cable channels she chose to watch, did not affect her supply of news.

While I focus on Facebook, this section's conclusions likely apply to other platforms personalizing content as well. For example, since at least 2016, Twitter has

[32] The control group participants subscribing to pro- and counter-attitudinal outlets are substantially different from each other. For example, among the 20 most popular liberal and conservatives pages, there is a difference of 0.32 standard deviations in the absolute value of ideology of subscribers to at least one pro- and counter-attitudinal outlet. Moreover, subscriptions to counter-attitudinal outlets occur several months later than subscriptions to pro-attitudinal outlets, and posts from more recent subscriptions are probably more likely to appear in the feed. The experiment assures that all subscriptions occur at the same time and due to a random offer.

been ranking tweets according to how interesting they would be for a user, based on factors such as the user's past relationship with the author. Therefore, it is plausible that tweets from pro-attitudinal accounts will receive a higher ranking. Furthermore, major news outlets have also started to personalize their websites and the articles they suggest to their customers.

VII. Conclusions

Consumption of news through social media is increasing, but the effect of social media on public opinion remains controversial. I show that news consumption on social media is an important phenomenon because consumers are exposed to different news on social media, individuals incidentally consume news when it becomes accessible in their feed, and exposure to news on social media affects attitudes.

This paper supports a more nuanced view regarding the effect of media on public opinion. On the one hand, I show that exposure to pro-attitudinal news increases affective polarization compared to counter-attitudinal news. This result provides a mechanism complementing other important studies finding that social media can increase polarization and raises concern since affective polarization may decrease trust in government and the accountability of elected officials. On the other hand, individuals are not easily persuaded by the political leaning of their news exposure. The results of the experiment are in line with the long-term increase in affective polarization, without an equivalent change in political opinions (Mason 2015). This suggests that a more segregated news environment may partially explain the increase in affective polarization over the past decades.

Methodologically, this paper has several limitations. First, I only observe online news consumption. While I show that the intervention did not have substantial spillovers across online outlets, to precisely measure total news consumption, future studies would need to collect consumption data from other mediums, such as television, as well. Furthermore, I collect data on browsing behavior and the Facebook feed on a computer, but a growing share of news is consumed through smartphones. Second, while I argue that due to the organic nature of the intervention, it is unlikely that experimenter effects play a major role in this study, I cannot rule out that the perceived expectations of the experimenter affected the results. Third, the endline survey suffers from high attrition. I use several methods to alleviate this concern, but attrition could still affect the survey outcomes. Fourth, the study does not generate random variation in exposure to moderate outlets and therefore cannot speak to their effects. Fifth, while the experiment has high external validity when it comes to analyzing partisan outlets on Facebook in 2018, the result may not hold for other periods. For example, Trump's presidency is exceptional in the stability of the president's approval ratings. If other opinions were relatively stable throughout the period as well, the null effect on political opinions could be explained by the period when the survey took place. Finally, I estimate all effects over several weeks or months, and the results may be different in the long term.

This study has important policy implications. I demonstrate that Facebook's algorithm limits exposure to counter-attitudinal news. Automated personalization of news content may have stronger impacts in the future, due to growth in online news consumption and advances in machine learning algorithms customizing news exposure. However, I also find that individuals are willing to engage with counter-attitudinal news. Therefore, even though social media platforms are associated with pro-attitudinal content, they can expose individuals to more perspectives. Suggestions include making algorithms more transparent, nudging users to diversify their feed, and modifying algorithms to encourage serendipitous encounters (Pariser 2011, Sunstein 2017). The experiment described in this paper essentially measures the effect of one such intervention and shows that a simple scalable nudge can effectively diversify news exposure and decrease polarization.

While social media algorithms may increase affective polarization through their effect on news consumption, platforms also have the potential to mitigate these effects.

REFERENCES

Allcott, Hunt, Luca Braghieri, Sarah Eichmeyer, and Matthew Gentzkow. 2020. "The Welfare Effects of Social Media." *American Economic Review* 110 (3): 629–76.

Bail, Christopher A., Lisa P. Argyle, Taylor W. Brown, John P. Bumpus, Haohan Chen, M. B. Fallin Hunzaker, Jaemin Lee, et al. 2018. "Exposure to Opposing Views on Social Media Can Increase Political Polarization." *Proceedings of the National Academy of Sciences of the United States of America* 115 (37): 9216–21.

Bakshy, Eytan, Solomon Messing, and Lada A. Adamic. 2015. "Exposure to Ideologically Diverse News and Opinion on Facebook." *Science* 348 (6239): 1130–32.

Baysan, Ceren. 2020. "Persistent Polarizing Effects of Persuasion: Experimental Evidence from Turkey." Unpublished.

Behaghel, Luc, Bruno Crepon, Marc Gurgand, and Thomas Le Barbanchon. 2015. "Please Call Again: Correcting Nonresponse Bias in Treatment Effect Models." *Review of Economics and Statistics* 97 (5): 1070–80.

Bennett, W. Lance, and Shanto Iyengar. 2008. "A New Era of Minimal Effects? The Changing Foundations of Political Communication." *Journal of Communication* 58 (4): 707–31.

Berinsky, Adam J., Gregory A. Huber, and Gabriel S. Lenz. 2012. "Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk." *Political Analysis* 20 (3): 351–68.

Boxell, Levi, Matthew Gentzkow, and Jesse M. Shapiro. 2018. "Greater Internet Use Is Not Associated with Faster Growth in Political Polarization among US Demographic Groups." *Proceedings of the National Academy of Sciences of the United States of America* 115 (3): 10612–17.

Bursztyn, Leonardo, Georgy Egorov, Ruben Enikolopov, and Maria Petrova. 2019. "Social Media and Xenophobia: Evidence from Russia." Unpublished.

Chen, Yuyu, and David Y. Yang. 2019. "The Impact of Media Censorship: 1984 or Brave New World?" *American Economic Review* 109 (6): 2294–332.

Chiang, Chun-Fang, and Brian Knight. 2011. "Media Bias and Influence: Evidence from Newspaper Endorsements." *Review of Economic Studies* 78 (3): 795–820.

Comscore. 2018. *Web Behavior Database Panel 2007, 2008, 2017, 2018.* Philadelphia: Wharton Research Data Services, University of Pennsylvania.

Coppock, Alexander, Emily Ekins, and David Kirby. 2018. "The Long-Lasting Effects of Newspaper Op-Eds on Public Opinion." *Quarterly Journal of Political Science* 13 (1): 59–87.

Davis, Mark H. 1980. "A Multidimensional Approach to Individual Differences in Empathy." *JSAS Catalog of Selected Documents in Psychology* 10.

DellaVigna, Stefano, and Ethan Kaplan. 2007. "The Fox News Effect: Media Bias and Voting." *Quarterly Journal of Economics* 122 (3): 1187–234.

De los Santos, Babur, Ali Hortaçsu, and Matthijs R. Wildenbeest. 2012. "Testing Models of Consumer Search Using Data on Web Browsing and Purchasing Behavior." *American Economic Review* 102 (6): 2955–80.

Durante, Ruben, Paolo Pinotti, and Andrea Tesei. 2019. "The Political Legacy of Entertainment TV." *American Economic Review* 109 (7): 2497–530.

Enikolopov, Ruben, Alexey Makarin, and Maria Petrova. 2020. "Social Media and Protest Participation: Evidence from Russia." *Econometrica* 88 (4): 1479–514.

Flaxman, Seth, Sharad Goel, and Justin M. Rao. 2016. "Filter Bubbles, Echo Chambers, and Online News Consumption." *Public Opinion Quarterly* 80 (S1): 298–320.

Garrett, R. Kelly, Shira Dvir Gvirsman, Benjamin K. Johnson, Yariv Tsfati, Rachel Neo, and Aysenur Dal. 2014. "Implications of Pro- and Counterattitudinal Information Exposure for Affective Polarization." *Human Communication Research* 40 (3): 309–32.

Gentzkow, Matthew, and Jesse M. Shapiro. 2010. "What Drives Media Slant? Evidence from U.S. Daily Newspapers." *Econometrica* 78 (1): 35–71.

Gentzkow, Matthew, and Jesse M. Shapiro. 2011. "Ideological Segregation Online and Offline." *Quarterly Journal of Economics* 126 (4): 1799–1839.

Gentzkow, Matthew, Jesse M. Shapiro, and Michael Sinkinson. 2011. "The Effect of Newspaper Entry and Exit on Electoral Politics." *American Economic Review* 101 (7): 2980–3018.

Gentzkow, Matthew, Jesse M. Shapiro, and Daniel F. Stone. 2015. "Media Bias in the Marketplace: Theory." In *Handbook of Media Economics*. Vol. 1B, edited by Simon Anderson, Joel Waldfogel, and David Strömberg, 623–45. Amsterdam: Elsevier.

Gerber, Alan S., Gregory A. Huber, and Ebonya Washington. 2010. "Party Affiliation, Partisanship, and Political Beliefs: A Field Experiment." *American Political Science Review* 104 (4): 720–44.

Gerber, Alan S., Dean Karlan, and Daniel 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.

Gosling, Samuel D., Peter J. Rentfrow, and William B. Swann. 2003. "A Very Brief Measure of the Big-Five Personality Domains." *Journal of Research in Personality* 37(6): 504–28.

Guess, Andrew. Forthcoming. "(Almost) Everything in Moderation: New Evidence on Americans' Online Media Diets." *American Journal of Political Science*.

Guess, Andrew, Brendan Nyhan, Benjamin Lyons, and Jason Reifler. 2018. *Avoiding the Echo Chamber about Echo Chambers.* Knight Foundation.

Guess, Andrew, Brendan Nyhan, and Jason Reifler. 2017. *"You're Fake News!" Findings from the Poynter Media Trust Survey.* The Poynter Ethics Summit.

Halberstam, Yosh, and Brian Knight. 2016. "Homophily, Group Size, and the Diffusion of Political Information in Social Networks: Evidence from Twitter." *Journal of Public Economics* 143: 73–88.

Hovland, Carl I. 1959. "Reconciling Conflicting Results Derived from Experimental and Survey Studies of Attitude Change." *American Psychologist* 14 (1): 8–17.

Iyengar, Shanto, and Masha Krupenkin. 2018. "The Strengthening of Partisan Affect." *Political Psychology* 39 (S1): 201–18.

Iyengar, Shanto, Yphtach Lelkes, Matthew Levendusky, Neil Malhotra, and Sean J. Westwood. 2019. "The Origins and Consequences of Affective Polarization in the United States." *Annual Review of Political Science* 22: 129–46.

Jo, Donghee. 2020. "Better the Devil You Know: Selective Exposure Alleviates Polarization in an Online Field Experiment." Unpublished.

Kennedy, Patrick J., and Andrea Prat. 2019. "Where Do People Get Their News?" *Economic Policy* 34 (97): 5–47.

Lee, David S. 2009. "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects." *Review of Economic Studies* 76 (3): 1071–102.

Lelkes, Yphtach, Gaurav Sood, and Shanto Iyengar. 2015. "The Hostile Audience: The Effect of Access to Broadband Internet on Partisan Affect." *American Journal of Political Science* 61 (1): 5–20.

Levendusky, Matthew. 2013. "Partisan Media Exposure and Attitudes toward the Opposition." *Political Communication* 30 (4): 565–81.

Levy, Ro'ee. 2021. "Replication Data for: Social Media, News Consumption, and Polarization: Evidence from a Field Experiment." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E121922V1.

Martin, Gregory J., and Ali Yurukoglu. 2017. "Bias in Cable News: Persuasion and Polarization." *American Economic Review* 107 (9): 2565–99.

Mason, Lilliana. 2015. "'I Disrespectfully Agree': The Differential Effects of Partisan Sorting on Social and Issue Polarization." *American Journal of Political Science* 59 (1): 128–45.

Mitchell, Amy, Jeffrey Gottfried, and Katerina Eva Matsa. 2015. *Millennials & Political News: Social Media—The Local TV for the Next Generation?* Pew Research Center.

Mosquera, Roberto, Mofioluwasademi Odunowo, Trent McNamara, Xiongfei Guo, and Ragan Petrie. 2020. "The Economic Effects of Facebook." *Experimental Economics* 23 (2): 575–602.

Müller, Karsten, and Carlo Schwarz. 2020. "From Hashtag to Hate Crime: Twitter and Anti-Minority Sentiment." Unpublished.

Okuyama, Yoko. 2020. "Toward Better Informed Decision-Making: The Impacts of a Mass Media Campaign on Women's Outcomes in Occupied Japan." Unpublished.

Orr, Lilla V., and Gregory A. Huber. 2020. "The Policy Basis of Measured Partisan Animosity in the United States." *American Journal of Political Science* 64 (3): 569–86.

Pariser, Eli. 2011. *The Filter Bubble: What the Internet is Hiding from You.* London: Penguin Press.
Parse.ly. 2018. "The Authority Report: 2018 Traffic Sources by Content Categories and Topics."
Peterson, Erik, Goel Sharad, and Shanto Iyengar. 2019. "Partisan Selective Exposure in Online News Consumption: Evidence from the 2016 Presidential Campaign." *Political Science Research and Methods* 1–17.
Rand, David G., Alexander Peysakhovich, Gordon T. Kraft-Todd, George E. Newman, Owen Wurzbacher, Martin A. Nowak, and Joshua D. Greene. 2014. "Social Heuristics Shape Intuitive Cooperation." *Nature Communications* 5: 1–12.
Reuters Institute. 2019. *Digital News Report 2019.* Oxford, UK: University of Oxford.
Strömberg, David. 2015. "Media and Politics." *Annual Review of Economics* (7): 173–205.
Sunstein, Cass. 2017. *#Republic: Divided Democracy in the Age of Social Media.* Princeton, NJ: Princeton University Press.
Wager, Stefan, and Susan Athey. 2018. "Estimation and Inference of Heterogeneous Treatment Effects Using Random Forests." *Journal of the American Statistical Association* 113 (523): 1228–42.
Williamson, Debra Aho. 2018. *US Time Spent with Social Media 2019.* eMarketer.
Yudkin, Daniel, Stephen Hawkins, and Tim Dixon. 2019. *The Perception Gap: How False Impressions Are Pulling Americans Apart.* More in Common.
Zhuravskaya, Ekaterina, Maria Petrova, and Ruben Enikolopov. 2020. "Political Effects of the Internet and Social Media." *Annual Review of Economics* 12: 415–38.