Social Influence and News Consumption*

Alex Moehring Carlos Molina

Job Market Paper

December 10, 2023

Click for most recent version

Abstract

Populations in several countries have become decidedly more polarized in recent decades. Many believe that social media, which facilitates interactions within echo chambers, is partly to blame. These interactions can trigger two distinct effects on the demand for biased news. First, individuals can be influenced by their peers' news consumption, for example, because they value keeping a news diet that is ideologically congruent with that of their peers. Second, individuals might purposefully skew their news consumption in anticipation that their peers will observe these choices. We design a field experiment on Twitter (renamed X in 2023) to separately identify the importance of both mechanisms. Our main result documents that, through these two mechanisms, online interactions with like-minded peers are not a major contributor to the demand for polarized news content. Our experiment induces variation in an individual's perceptions of the political leanings of their peers' news consumption and the visibility of their own news consumption to their social media followers. We track participants' sharing behavior and news consumption, proxied by the news outlets they follow. We find no evidence to support the first channel: our experimental variation influences respondents' beliefs about the news diets of their peers, but they do not respond by changing their own news diets. In contrast, we find that participants alter their news diet considerably when they believe their peers will observe these choices, as in the second channel. Interestingly, individuals primarily wish to present themselves as following a balanced set of news. Therefore, our paper uncovers one mechanism through which social media can attenuate the demand for polarizing content: as these platforms amplify the visibility of user interactions, which increases the importance of social image concerns, users adjust their news consumption to be more balanced.

Keywords: biased news, social image, learning, polarization.

JEL Classification: C93, D72, D83, L82.

---

*Moehring: Ph.D. candidate, MIT Sloan, `moehring@mit.edu`. Molina: Ph.D. candidate, MIT Economics, `camolina@mit.edu`. We are extremely grateful for the generous advice and support of Daron Acemoglu, Ben Olken, and Frank Schilbach. For helpful feedback, we thank Nicolás Ajzenman, Cevat Giray Aksoy, Ceren Baysan, Dean Eckles, Leopoldo Fergusson, Martin Fiszbein, Kelley Friel, Alexey Makarin, Jacob Moscona, Catherine Tucker, MacKenzie Walker as well as seminar participants at MIT. This work received support from the George and Obie Shultz Fund at MIT, the Institute for Humane Studies (IHS017300), and the Bradley and Hewlett Foundations. IRB approval for this project was obtained from the Massachusetts Institute of Technology Committee on the Use of Humans as Experimental Subjects (protocols E-3516 and E-4465) and the experiment and primary analysis were pre-registered at the American Economic Association's registry for randomized controlled trials (protocol AEARCTR-0011147), `https://www.socialscienceregistry.org/trials/11147`.

1 Introduction

Social media increasingly mediates exposure to news content. In 2021, 30% of citizens around the world and nearly half of Americans reported consuming news through social media.[1] This represents a dramatic shift over the last 20 years before the creation of such platforms. As individuals come to increasingly rely on social media for information, there are growing concerns that such platforms play an important role in polarizing individual beliefs. A leading hypothesis regarding how social media platforms affect polarization is that they facilitate interactions within echo chambers – where individuals are more likely to be exposed to content from peers with like-minded beliefs – thus limiting users' access to counter-attitudinal opinions [Settle, 2018, Sunstein, 2018]. Understanding how interactions with peers affect users' preferences for politically biased content is critical to assessing how echo chambers – and ultimately social media – affect polarization.

We zero in on two distinct channels through which social influence can impact preferences for biased news. The first is the *peer effects channel*: peers directly influence individuals by following particular news outlets. For example, users may seek to consume a news diet that is ideologically congruent with their peers.[2] The second is the *social image concerns channel*: individuals might purposefully alter the news outlets they consume when they expect peers to observe their choices.[3] The extent to which these channels affect a person's demand for polarizing content depends on how much they trigger the person's need to resemble their peers (or digress from them) as well as the person's preferred news ideology relative to that of their peers.[4]

There is little causal evidence of the importance of these two channels of social influence in determining users' preferences for biased news partly due to the challenges associated

[1]These figures are based on the Pew Research Center 2021 Media Consumption Survey and the Reuters Institute's 2021 Digital News Report. See also Mason Walker and Katerina Eva Matsa, "News Consumption Across Social Media in 2021," September 20, 2021, `https://www.pewresearch.org/journalism/2021/09/20/news-consumption-across-social-media-in-2021/`.

[2]This channel is motivated by and closely related to the literature studying the impact of social norms on economic behaviors ranging from household finance [Lindbeck, 1997], labor force participation [Bursztyn et al., 2020b], and extremism [Bursztyn et al., 2020a]. The peer effects channel of social influence investigates the importance of descriptive norms on preferences for biased news.

[3]This channel is motivated by a growing literature in economics that studies the importance of social image concerns in driving economic behaviors including political actions [DellaVigna et al., 2016], educational investments [Fryer Jr and Torelli, 2010, Bursztyn and Jensen, 2015], and labor decisions [Bursztyn et al., 2017]

[4]For instance, these channels would lead to more (less) demand for biased content for a focal user in the specific case where these channels stimulate conformity –the need to resemble their peers– and the focal user has peers with more (less) extreme preferences for biased news.

with identifying the causal effect of peers' behavior on a focal user. These challenges include individuals forming their social networks endogenously as well as the reverse causality problem: a focal user can influence their peers *and* be influenced by them. The need to separately estimate the importance of these two distinct mechanisms further complicates the identification process, which involves inducing variation in (1) a user's beliefs about the news diets of their peers and (2) how visible a user's news consumption is to their peers. It is also often difficult to observe what news users consume – providing an additional challenge to understanding the extent to which social influence shapes preferences for biased news.

We design a novel field experiment on Twitter[5] to separately identify the importance of the peer effects and social image concern mechanisms. We recruited a sample of 3,757 users via Twitter ads between March and June 2023. To overcome the challenges of separately identifying these channels, we induce variation in both participants' perceptions of the political leanings of their peers' news consumption and the visibility of their own news preferences to their social media followers. We observe whether participants choose to publicize their news diet to their followers and/or alter their news consumption – which we measure as changes to the news outlets they follow on Twitter. We test whether information regarding the news consumption of their peers affects participants' news consumption (peer effects) by tracking how our randomizations influence their beliefs about the news diets of their peers and the subsequent news choices that participants make. Additionally, we evaluate the presence of strategic manipulation of social media behavior in response to social image concerns by testing whether participants modify which news outlets they follow depending on whether their peers can observe these choices.

We share with each participant a summary of the ideological composition of their news diet. This summary contains both the quantity of news outlets a participant follows in addition to the average ideological slant of the news outlets they follow (how preferred an outlet is by Democrats vs. Republicans as described in Robertson et al. [2018]). We then cross-randomized participants into two experimental conditions. Participants in the *peer information* condition received an additional summary of the ideological composition of their peers' news diets. In the *disclosure* condition, participants in the treatment group are incentivized to reveal to their peers – via a tweet – a summary of the political ideology of the news they consume. The control group has an incentive to share a placebo message that contains no private information.[6] Immediately after we sent individuals in both groups

[5] In this paper we refer to this platform as 'Twitter' (renamed 'X' in 2023) and posts and reposts as 'tweets' and 'retweets' as they were known at the time of our experiment.

[6] To incentives these tasks, participants in both groups are offered the possibility to participate in a

information about this incentive, but before they are asked to tweet the message, they receive information explaining how different news outlets can change the ideological slant of the news to which they are exposed.

Our main result is to document that, through these two mechanisms, online interactions with like-minded peers are not a major contributor to the demand for polarized news content. Importantly, we find that participants tailor their news diet when they believe their news diets will be observed by their peers, which supports the social image concerns channel. Ex ante, it is unclear whether we should expect participants to make their news diets more moderate or more ideologically congruent with their peers if they believe their peers will observe the news outlets they follow. Interestingly, we find that participants in the disclosure treatment group make their news diets more moderate, moving toward the center of the ideological spectrum. Moreover, we find little evidence of the importance of peer effects. Although our treatment variation in the peer information condition induce changes in the posterior beliefs that participants have about their peers' ideological news consumption, this information does not translate into individuals making changes to their news choices.

Comparing the treatment and control groups in the disclosure condition implies the following five results. First, the incentive to publicly share information in the disclosure condition has the expected effect on sharing behavior. The treatment group is 8.2 percentage points (standard error of 0.8) more likely to share the summary of the political ideology of their news diet with their peers. We randomize the expected payoff that the control and treatment groups have for tweeting the placebo message and the news diet summary respectively. We find that compliance is greater when the expected payoff increases. Despite facing the same monetary incentives, the treatment group is over three times less likely than the control group to comply with their respective task, consistent with the idea that there is in fact a social image cost of making the news diet summary public.[7]

Second, the disclosure treatment has a large impact on the news outlets participants follow on Twitter. Those who are incentivized to reveal information to their peers are more likely to change which news sources they follow. The treatment increased the probability that a participant would make at least one change to their news diet by 21.4% (standard error of 4.6%) and boosted the total number of outlets followed by 28.6% (standard error

lottery for a randomized amount if they comply.

[7]Note that this cost goes beyond the simple act of tweeting, which can have a cost itself, as the control group is also incentivized to tweet the placebo message.

of 6.0%) relative to the control mean. We find that the magnitude of the change in the average slant of outlets a user follows was 33.7% (standard error of 10.5%) larger than the control mean.

Third, on average, participants in the treatment group are 40.9% more likely to shift their news diet toward the ideological center and 31.6% more likely to move it toward their peers (standard errors of 8.2% and 7.8%, respectively). There is important heterogeneity, however, based on the initial position of the participant's news diet on the ideological scale and the position of the average news diet of their peers. In our sample, around 80% of participants follow at least one news outlet at baseline. We find that these "news consumers" move strongly toward the peer/center when both are in the same direction, and still move toward the center when they are in opposite directions. The remaining 20% of "non-news consumers" in the sample tends to move toward their peers (although these changes are typically of a much smaller magnitude); if their peers are also not politically engaged, the treatment has no detectable effect. Therefore, it appears that users predominantly want to be perceived as having a neutral news diet.

Fourth, we find that the effects of the disclosure treatment are stable over time and persist for months after the experiment. We test this by periodically collecting public data on the participants after the intervention using the Twitter API. The differences between the treatment and control groups in the number of outlets followed and their average slant persists over time. This finding is not a mechanical result from users not adjusting which outlets they follow, since participants in the control group frequently change the outlets they follow over time.

Fifth, we find that the disclosure treatment increases engagement with news outlets in the form of retweets, likes, and tweets mentioning a news source. This effect is specific to news outlets, as the treatment does not impact engagement with non-news Twitter accounts. Moreover, we find that most of the effect on engagement with news sources is driven by outlets followed during and before the experiment. The fact that the disclosure treatment has important effects on long-term news diets and engagement rejects the hypothesis that users in this treatment merely adjust their news diet temporarily and revert to their original news sources afterwards.

The peer effect channel explores how learning about which news outlets their peers follow shapes participants' own news diets. Our experimental manipulations *do* change participants' beliefs about which news sources their peers follow: those in the peer treatment group are 32.3% (standard error of 6.2%) more likely to update their beliefs about their peers relative to the control mean. Individuals mostly update in the direction of the feedback (relative to prior beliefs). They are more responsive when the information they receive is further from their prior beliefs, and the treatment has a negligible effect when individuals hold accurate beliefs about their peers.

Although the treatment successfully induces variation in participants' beliefs about their peers, we find no evidence that this updating of beliefs leads to changes in news diets. The treatment group did not differentially change the number or slant of outlets followed. These effects are statistically indistinguishable from zero and small in magnitude. Our main estimates indicate that individuals in the treatment group adjust the slant of their news diets (in the opposite direction of the feedback) by 0.01 standard deviations (standard error of 0.02) relative to the control group. This is despite the treatment having a significant effect on participants' beliefs about their peers, as the treatment causes users to update their beliefs about the news diets of their peers in the direction of the feedback by 0.12 standard deviations (standard error of 0.02).

As complementary evidence, we exploit a different source of variation to estimate the importance of peer effects. We provide participants in the treatment group with an unbiased but noisy estimate of the slant of peers' news diets based on a random sample of peers due to an API rate limit that prohibits collecting more exhaustive data in real time. The noise component in the feedback generates additional exogenous variation that we use to identify how left vs. right shocks (conditional on the peers' true positions) affect participants' beliefs about the news diets of their peers. We find that individuals who receive right-leaning shocks about which news sources their peers follow update their posterior about the position of their peers' news diets to the right by 0.24 standard deviations (standard error of 0.02) relative to those who receive left-leaning shocks. We again find that this effect is greater when participants receive a larger shock to their beliefs about their peers' favored news sources, and dissipates when sent a signal that is close to the true position of their peers. Although people's beliefs about their peers change, we again find that our treatment has negligible effects on the composition of participants' news diets using this alternative strategy: those who receive a signal about their peers' ideology being to the right (left), choose news outlets that have a slant of 0.004 standard deviations (standard error of 0.02) to the left (right). This coefficient is not statistically significant and is small compared to the magnitude of the effect on beliefs. We also conduct a placebo exercise in which we estimate this regression across the control group, thus relying on variation in the feedback they would have received had they been assigned to the treatment group. Reassuringly, this placebo has no significant effects on either belief updates or news consumption choices.

Overall, this paper investigates the role of two important mechanisms that explain individuals' preferences for biased news – social image concern and peer effects. We find that individuals significantly adjust their news consumption choices when they believe their peers are observing these choices, which suggests the social image concerns mechanism is an important channel. If anything, this channel *reduces* the demand for biased news, as individuals want to conceal their interest in extreme news choices. We find little support for the peer effects channel: changes in beliefs about the political ideology of the news that peers consume have little impact on determining individuals' own news consumption choices.

This paper contributes to several strands of the existing literature. Our study is related to prior research on the determinants of individuals' preferences for news outlets, including theoretical work explaining preferences for politically biased or slanted news [Suen, 2004, Gentzkow and Shapiro, 2006, Mullainathan and Shleifer, 2005]. Gentzkow and Shapiro [2010] and Chopra et al. [2022, 2023] have documented these preferences empirically. Our paper advances empirical studies of how a peer's news consumption impacts an individual's own consumption [Messing and Westwood, 2014, Aral and Zhao, 2019, Nyhan et al., 2023] by investigating a novel channel that could induce individuals to prefer politically slanted news – social image concerns. Our results demonstrate that social influence plays an important role in discouraging the consumption of biased news, mainly due to social image concerns (individuals change their behavior to influence how they present themselves to their peers). We find that the peer effects channel does not play a significant role in influencing news diets: our treatments influence participants' beliefs about which news outlets their peers follow, but they do not respond by changing their own news diets.

Our results shed light on the role that information technologies, particularly social media, play in explaining preferences for biased news and ultimately shaping political attitudes [Gentzkow and Shapiro, 2011, Enikolopov et al., 2020, Boxell et al., 2022]. Much of the recent literature investigates the causal effect of exposure to pro-attitudinal (as opposed to counter-attitudinal) news content on polarizing attitudes (based on the idea that social media algorithms expose users to more pro-attitudinal content to maximize engagement) [Guess and Coppock, 2020, Levy, 2021, Broockman and Kalla, 2022, Casas et al., 2022]. Our paper contributes to this literature by focusing on a different channel: that social media influences preferences for biased news (and thus polarization) by facilitating interactions within echo chambers, where users are more likely to interact with peers who share similar ideological beliefs. Our paper is among the first to provide causal evidence that social influence on social media drives preferences for politically slanted news. Contrary to popular beliefs, we find little evidence that social influence causes users to consume more politically slanted news outlets. Instead, our paper reveals a mechanism that could mitigate users' preferences for politically slanted outlets. As social media platforms amplify the visibility of user interactions, concerns about social image concerns increase. As a result, our findings suggest that users skew their news consumption to be more moderate to signal a preference for a balanced news diet to their peers.

Our paper advances economics research on the effects of social image concerns [Bursztyn and Jensen, 2017], particularly how they relate to political attitudes [Gerber et al., 2008, Funk, 2010, DellaVigna et al., 2016]. We provide the first evidence on how social image concerns drive the demand for (un)biased news. Unlike most papers in this literature, we study these concerns in the context of social media, and connect recent psychology studies that highlight the important role of social media in exacerbating image concerns and the understudied results of this effect on online behavior [Fardouly and Vartanian, 2016].

The rest of the paper is organized as follows. The next section describes the experimental design, empirical strategies, outcome variables of interest, and other empirical setups of the field experiment. Section 3 presents the basic statistics of our sample. Sections 4 and 5 describe the main results regarding the role of social image concerns and peer information in shaping individual preferences for biased news, respectively. Section 6 concludes, while the (online) Appendix contains several additional descriptions and robustness checks.

2 Experimental Design and Data

2.1 Experiment Overview

Figure 1 summarizes the design of the experiment. We used Twitter ads to recruit American adults between March and June 2023. A total of 954,913 unique users saw the ad, of whom 12,940 clicked on it.[8] Users who clicked on the ads were directed to the survey landing page, which contained an overview of the experiment, discussed the incentives to participate

[8]Appendix Section A.1 contains details on the ads. This 1.36% click-through rate is slightly higher than the average rate for Twitter ads across all industries (0.86%) (see `https://www.brafton.com/blog/social-media/social-advertising-benchmarks/`) and comparable to other studies in the literature reporting similar metrics in Facebook (Allcott et al. [2020] and Allcott et al. [2022] report a click-through rate of 1.7% and 0.8%, respectively).

in the study, requested the participants' Twitter handle or username, and presented the consent form. A total of 5,192 individuals who consented to participate and had a public Twitter account were invited to begin the baseline survey, which recorded demographics and a wide range of pre-treatment covariates including political engagement, ideology (of the user and their peers), political knowledge, and affective polarization. It also elicited self-reported beliefs about the political bias of the news outlets that participants were following on Twitter, as well as those followed by their peers. Our randomization includes the 4,548 participants who completed the baseline survey.

While participants were completing this baseline survey, we used the Twitter API to scrape the participant's network in real time – the list of accounts the participant followed on Twitter and the list of accounts that followed the participant. We also scraped the network of a random sample of five of the participant's followers. We use all of this scraped data to construct our treatments.[9]

As illustrated in Figure 1, we cross-randomized participants into either the peer information or disclosure condition. After completing the baseline survey, participants received a summary of their news diet that details which news outlets they follow on Twitter. We constructed this summary using the data that we scraped for each user – the number of news outlets they follow (relative to the average Twitter user) and the average political slant of these outlets. While everyone had access to the summary information of their own news diet, only the treatment group in the peer information condition (the first randomization) also received information on their peers. This information constituted an estimate of the average slant of five randomly chosen followers (see footnote 9). Appendix Figure A-1 shows the design of the infographics we used to convey this information. After participants reviewed this information, the second randomization assigned users to either the disclosure treatment or the control group.

Within the disclosure condition, participants were randomly assigned to either a control or treatment group. Those in the treatment group were incentivized to tweet an infographic revealing the summary of their news diet. Importantly, the instructions clearly stated that participants would not be asked to share the summary until after they had had the chance to modify which news outlets they follow. The control group received a similar message but was instead incentivized to tweet a referral link to promote the study; thus, there was

[9] We only retrieved the network of up to five followers mainly due to restrictions on the number of queries that can be made to the Twitter API in a short period of time. We discuss this in more detail in Section 2.4 and show how we use this limitation to leverage additional variation to study the role of peer information.

no incentive to disclose any personal information.

Once participants were informed of the incentives and required tasks, they continued to a page where they were given a chance to change the news outlets they follow. We provided information on three conservative and three liberal news outlets that participants did not previously follow and explained to them the impact of following any of these news sources on their average political slant. Appendix Figure A-2 shows an example of this information. Subjects were allowed to follow as many of the recommended outlets as they wanted in addition to following (or unfollowing) other news outlets.

After participants had an opportunity to change which news outlets they follow, we re-scraped the network of their Twitter accounts. We then presented them with an updated news diet summary. All participants were then given the option to tweet the referral link with this updated summary. All participants had the option to select one of two buttons – "Share Referral Link" or "Share News Diet Summary" – that would automatically draft a tweet for them to review and share. No mention of the incentive or any other differential information was provided at this stage. Appendix Figure A-3 shows an example of the drafted tweets that participants were shown if they clicked on these buttons.

In a last step, users were invited to complete a final questionnaire that again elicited beliefs about ideology, the political bias of the news outlets they follow on Twitter, and the political bias of the news outlets followed by their peers. In the months after the intervention, we tracked the activity of our participants and their peers on Twitter. This includes both changes in the user's network (and therefore news diets) and the participant's engagement through tweets, likes, and retweets.

2.2 Incentives to Participate

The consent page offered individuals the chance to enter a lottery to win $200 upon completion of the endline survey. Participants were also told they could be selected during the survey to complete an additional task, which would enter them into a bonus lottery (more details below). There were no other financial incentives to participate in the study. We also appealed to participants' sense of altruism, emphasizing that their participation would help our careers as junior researchers. Moreover, they were told that by completing the survey, they could learn about the political bias of the news to which they were regularly exposed on Twitter.

We used the bonus lottery to incentivize participants to tweet either the news diet summary (if selected to the public group) or the referral link with no personal information (if selected to the private group). Regardless of their treatment status, we informed all participants that they would be eligible to participate in the bonus lottery if they completed the task associated with their treatment group. To estimate the cost of disclosing this information, we randomized the amount of the bonus lottery to either $100 or $200. In general, we made clear to participants that the probability of winning each lottery was independent of the other lotteries. That is, compliance with the bonus task did not impact the likelihood of winning the main lottery upon completion of the survey.

We elicited participants' beliefs about their chances of winning both of these lotteries in the endline survey and found that the median participant believed there was a total expected monetary reward of $2.50 for finishing the study. Approximately $2 of this was due to beliefs about the main lottery for which participants were eligible upon completing the endline survey. In addition, 17% of participants who complied with the bonus task placed an expected value of $3 in the bonus lottery.

2.3 News Outlets Dataset

We constructed a dataset of 1,170 US news outlets on Twitter that we used to assemble the infographics conveying the participants' news diet summary (see Appendix Figure A-1). We focus on sources that predominantly cover hard news (e.g. significant political, economic, and societal developments) rather than soft news (e.g. entertainment, sports). Therefore, we use the terms "outlets" and "publishers" interchangeably to refer to these hard news publishers.

For every publisher in our dataset, we assigned a publisher slant score from -1 and 1 that represents its relative propensity to be shared on Twitter by Democrats relative to Republicans based on Robertson et al. [2018]. An outlet only shared by Democrats has a slant of -1, a moderate news outlet (shared by both) has a slant of approximately 0, and one only shared by Republicans has a slant of 1. The primary advantage of this scoring system over other methods of assessing publisher bias is its extensive coverage (the dataset provides ratings for more than 19,000 domains). Robertson et al. [2018] demonstrate that this measure is consistent with other expert, crowd-sourced, audience-based ratings that typically have less coverage.[10]

[10] The correlation between their score and the Bakshy et al. [2015] score (which is based on the self-reported ideology of Facebook users sharing articles from the domains), a measure commonly used in the literature reporting ratings for 500 news domains, is 0.96.

Appendix Section A.3 discusses this measure and the methodological methods we used to identify news outlets from Robertson et al. [2018]'s list of domains and map these publishers to Twitter accounts. Appendix Figure A-4 displays the distribution of the slant scores of our final dataset of news outlets.

2.4 Outcomes Variables and Empirical Strategy

We designed our experiment to study the effect of the interventions on three complementary sets of outcomes. The first is changes in participants' beliefs about their own and their network's consumption of biased news based on the endline survey. The second set of outcomes is changes in participants' long-term choices of news outlets to follow on the platform based on changes to the outlets they followed during the experiment. The third is changes in engagement with news outlets based on the engagement behavior collected on Twitter after the experiment.

To estimate the effects of social image concerns and peer information, we compare the treatment and control groups in each of our randomizations using the following intention to treat (ITT) regression:

$$Y_i = \alpha + \beta D_i + \varepsilon_i \tag{1}$$

where Y_i is the outcome of interest and D_i is a dummy that equals 1 if the participant receives the relevant treatment. Because of our cross-randomized setting, the coefficient of interest β in equation (1) should be thought of as the ITT effect across both types of participants: those who received the other treatment and those who did not [Muralidharan et al., 2023]. For the social image concern treatment, for instance, β is the effect of the public treatment across participants who received peer information and those who did not. We use this approach for easiness of exposition, but in Appendix Section A.4 we discuss the results when estimating ITT effects for non-mixed groups using a fully saturated model that includes both treatments and their interaction. The interaction between both treatments is non-significant and small in magnitude, which indicates that the estimates using both methods are very similar.

We designed our experiment to exploit an additional source of variation when studying the role of peer information. As stated in Section 2.1, participants in the peer information treatment are exposed to an estimate of the average slant of a random sample of five of their followers – an unbiased but noisy estimate of the true slant of their peers' news diets. After the participants completed the endline survey, we continued to scrape the network of their followers to compute a more precise estimate based on this larger number of followers. Due to API restrictions, we were unable to scrape the network of all followers; we therefore used the average of a random sample of approximately 20 followers. We exploit the difference between the (noisy) feedback that participants received ($S^F_{feedback}$) and the more precise measure based on the larger number of followers (S^F_{truth}) as an instrument for participants' beliefs about the ideological positions of their followers. The noise component is crucial as it implies that participants with identical followers can be shown different estimates of the slant of their peers' news diets due to random sampling that is orthogonal to pre-treatment characteristics. This allows us to causally identify the effect of the noisy feedback on posterior beliefs and news choices using the following regression:

$$Y_i = \alpha + \gamma \mathbb{I}[S^F_{feedback} > S^F_{truth}] + \delta X + u_i \tag{2}$$

where $\mathbb{I}[c]$ is a dummy indicator if condition c holds and γ is the coefficient of interest. X represents the controls we include in the regression, which include the accurate signal (S^F_{truth}) as well as the individual's prior beliefs. Unless specified otherwise, standard errors in all our specifications are robust to heteroscedasticity.

2.5 Preanalysis Plan

This experiment and the primary analyses were pre-registered in the AEA RCT registry.[11] However, the experiment deviated from the pre-analysis plan in two ways. First, we have not reported the results of an analysis that instruments a user's decision to share their news diet summary with the treatment status of the disclosure randomization, which would have allowed us to study the causal effect of a user sharing the summary on the participant's own news diet. We did not pursue this analysis because we observed that users in the disclosure treatment who did *not* share their news diet summary were also more likely to make changes to which news outlets they follow than participants in the non-disclosure condition. This could be because users who intended to share their news diet summary made changes to which news sources they follow and then decided not to share. Therefore, we interpret the treatment effect of the incentive to share as capturing the effect on both (1) users who eventually shared and (2) users who intended to share but did not follow through. In this case, the instrumental variables estimates are difficult to interpret and

[11] See `https://www.socialscienceregistry.org/trials/11147` for the pre-registration (protocol AEARCTR-0011147).

therefore have not been pursued. In a second deviation from the pre-analysis plan, we pre-registered that we would flexibly include pre-treatment covariates in the treatment effect estimates, a task that has not yet been pursued.

3 Descriptive Statistics

In this section, we present the basic summary statistics of our sample. Of the 4,548 participants who completed the baseline survey, our final sample contains the 3,757 participants who completed the endline survey – an attrition rate of 17%. Appendix Figure A-7 shows that the differential attrition rates across the two randomization conditions were relatively small (less than 2%) and are not statistically significant.

Table 1 quantifies the representativeness of our sample on observables by comparing the demographics of our impact evaluation sample to the US adult population. Our sample is predominately white, more educated, more heavily male, and older than the US adult population. Appendix Section A.5 discusses the robustness of our results when we use sample weights to adjust for these observable differences and demonstrates that the treatment and control groups are balanced on observables.

Appendix Table A-1 reports the descriptive statistics of our sample in terms of engagement. The median participant in our sample tweets around four times every five days, has 148 followers, and follows 500 accounts. There are significant differences at both ends of the engagement distribution: the median participant in the top 10% of tweeters sends 56 times more tweets than the median participant in the bottom 10% of tweeters. In general, these figures indicate that our sample is comparable to, and if anything more engaged than, a nationally representative panel of US adults with an active Twitter account.

Appendix Figure A-8 displays the distributions of the number of news outlets users follow and the average slant of these outlets (conditional on following at least one news outlet). We report these distributions across three samples: the study participants, the participants' peers, and a random sample of Twitter users.[12] The median participant in our sample follows four news outlets and is much more politically engaged than both their peers and the random sample of Twitter users. Around 20% of participants do not follow any publishers, which is substantially lower than the 38% or 55% for the participant's peers and random sample, respectively. Participants who do not follow any news outlets are classified as having a neutral news diet and are therefore assigned a slant score of zero.

[12] Appendix A.2 explains how we constructed the random sample of Twitter users.

When presented with the news diet summary (e.g. Appendix Figure A-1), users who do not follow any news outlets are depicted at the center of the political scale and at the bottom of the distribution for the number of followed outlets. We discuss differential effects for this subgroup when presenting the main results.

We compare the summaries of participants' news diets constructed during the experiment (which we show participants in the peer information condition, see Section 2.1) to their self-reported pre-experiment values to assess the reliability of the summaries we presented participants. Appendix Figure A-5 plots the average slant of participant news diets based on their self-reported ideology and the self-reported slant of their news diet; the news diet summary we construct during the experiment is highly correlated with both self-reported measures. This strong correlation increases our confidence that the summaries capture important features of participants' news diets. Moreover, we find that participants' self-reported beliefs about the ideology and news diets of both their followers and who they follow on Twitter is highly correlated with the average slant of publishers followed by their followers. This analysis suggests that the measure we show participants reflects, on average, participants' own beliefs about their peers on the platform.

4 The Role of Social Image Concerns

This section presents the treatment effect estimates of the disclosure treatment following equation (1). These estimates capture the role that social image concerns play in determining which publishers a participant chooses to follow. We exploit the difference between the treatment and control groups in the disclosure condition discussed in Section 2.1 to estimate the causal effect of the incentive to disclose the news summary on the various outcomes of interest. Importantly, the effects are not driven by receiving any incentive to tweet, as users in both the treatment and control groups receive an incentive to send a tweet. Only the *content* of the tweet they are incentivized to share varies.

4.1 Do Incentives to Share Information Impact Disclosure?

Figure 2 indicates that our treatment had the intended effects on compliance. The two panels illustrate the fraction of participants that tweeted the referral link and news diet infographic for the control and treatment groups – 16.6% and 7.3%, respectively. The control group was 9.3 pp (standard error of 1.1) more likely to tweet the referral link. However, 10.5% of the participants in the treatment group tweeted the news diet infographic, compared to only 2.3% in the control group. Thus the treatment group was 8.2 pp (standard error of 0.8) more likely than the control group to tweet the referral link. Both differences are statistically significant and in line with the incentives offered to each group.

4.2 Do Social Image Concerns Affect News Choices?

One of our main research questions is how social image concerns influence which news outlets an individual chooses to follow on Twitter. In this section, we focus on outcomes related to individuals' behavior during the experiment. These outcomes include changes made on the recommendation page after being exposed to information about different news outlets and their potential effect on participants' average news diet, as well as changes participants made outside the experiment platform, such as unfollowing or following publishers not recommended to them.

The three panels in Figure 3 show, from left to right, the probability of participants making at least one change to their news diet, the change in the total number of outlets followed, and the absolute value of the change in the average slant of the outlets followed relative to those followed before the randomization, respectively. We find that a high fraction of participants in the control group made substantial changes to their news diet summary: 34% of those in the control group made at least one change; those who made changes followed an additional 1.9 publishers on average.

More importantly, we find that participants in the treatment group made even more dramatic changes to their news diet summaries. Relative to the control mean, the treatment increased the probability of making at least one change to the news diet by 21.4% (standard error of 4.6%), the total number of outlets followed by 28.6% (standard error of 6.0%), and the absolute value of the change in the slant of the outlets followed by 33.7% (standard error of 10.5%).

4.3 What Factors Drive Image Concerns?

The results in the previous section indicate that social image concerns play an important role in determining which publishers a participant chooses to follow. In this section, we seek to better understand how participants would like to be perceived by their peers. We investigate two potential mechanisms. First, individuals may be wary of revealing the ideological position of the news outlets they follow. Some may wish to be perceived as balanced and unbiased, forming opinions based on a broader spectrum of information. Others might prefer to exhibit more extreme positions due to a desire for ideological reinforcement and to demonstrate their strong convictions.

In a second potential mechanism, individuals might be concerned about their relative standing within their network. An individual could prefer to resemble their peers, for instance, if she is afraid of peer reactions when deviating from a news diet aligned with their ideological position. Alternatively, an individual could pursue a strategy of "digressing to impress" if, for instance, she believes that engaging with cross-cutting content can signal to peers her awareness of diverse perspectives, thereby enhancing her perceived intelligence and credibility in social interactions. To summarize, it is ex ante unclear whether social image concerns will cause participants to adjust their news diets toward their friends, toward more neutral publishers, or away from their friends.[13]

We test both of these hypotheses by investigating whether people choose news outlets that, on average, move their news diet toward (or away from) the ideological center and their followers. We focus on the group of news consumers, which we define as participants who follow at least one news outlet at baseline (around 80% of our sample), as the slant and moving toward the center are not well-defined variables for the sample of non-news consumers, as we discuss below. As depicted in Figure 4, we find that compared to the control mean, participants in the treatment group are 7.68 and 6.25 pp (standard errors of 1.54 and 1.55, respectively) more likely to adjust their news diet toward the center and their peers, respectively.

To better understand how users would like to be perceived, we estimate the heterogeneous treatment effect of the disclosure treatment on moving toward either a participant's peers or the center, conditioning on the relative position of the participant's news diet, their peers, and the center. Figure 5 presents the results of this analysis. The top panel displays the treatment effect of the disclosure treatment among users whose peers are toward the center. These represent Democrats (Republicans) whose peers are to their right (left). In this case, movements toward the center and peers are identical and there is a strong effect: 8.0% of users adjust their summary statistic toward both the center and their peers. The bottom panel displays the treatment effect among users whose peers and the center are in opposite directions: Democrats (Republicans) whose peers are to their left (right). In this case, the disclosure treatment induces an additional 6.4% of users to move toward the

[13]These need not be mutually exclusive, as the treatment can also induce variation in the probability of making changes.

center and 2.8% fewer participants to move toward their peers. The negative treatment effect on moving toward peers, though not statistically different from 0, implies that the disclosure treatment induces some users who adjust their news diets toward their peers in the control group *not* to change their news diet or move toward the center.

Appendix Figure A-9 reports the results for non-news consumers. The top panel displays the treatment effect among users who do not follow any news outlets at baseline and whose peers also do not follow any news publishers. We find that these users are highly unlikely to follow any news outlets in either the control or treatment group: 87% of users in both groups do not follow any news outlets. The bottom panel illustrates the treatment effect among users who do not follow any news outlets at baseline but whose peers do follow news outlets. These users are more likely to adjust their news diet toward that of their peers, although they do so to a much lesser extent.[14]

Together, these findings suggest that users prefer to move toward both their peers and the center. When a user's peers are toward the center, users are more likely to adjust their summary statistic toward the center (and their peers) than when the two are in opposite directions. When a user's peers are *away* from the center, users tend to move toward the center rather than toward their peers. This suggests that when the center and a user's peers are in conflict, the desire to adjust their summary toward the center to be perceived as more moderate dominates. Finally, users who do not follow any publishers at baseline tend to follow publishers that are closer to their peers. Therefore, it appears that users predominantly want to be perceived as having a neutral news diet; to a lesser extent, they also care about following news outlets with a similar slant to those of their peers.

4.4 Persistence of the Treatment Effects

The treatment effects discussed above demonstrate that incentivizing participants to share their news diet summary induces them to change the news outlets they follow during the experiment. Here, we investigate whether these changes are short-lived or if they persist after the experiment by re-estimating equation 1 for outcomes that are observed periodically after the experiment, again using Twitter API data. We estimate the effects of the disclosure treatment on the probability of making at least one change to the news diet, the change in the total number of outlets followed, and the absolute value of the change in

[14] The disclosure treatment increases the absolute value of the slant change by 7.78% of a standard deviation (standard error of 1.85%) in the sample of news consumers, but by only 2.18% (standard error of 3.64%) for non-news consumers with news consumer peers.

the slant of the outlets followed (relative to the outlets participants were following before the randomization). Figure 6 displays the results. The treatment effects observed during the experiment are quite persistent: there is little decline over time. For all outcomes, we cannot rule out the possibility that the treatment effect after the end of the experiment is statistically different from that measured during the experiment.

4.5 Impact on Engagement with Publishers

In this section, we test the extent to which participants engage with news outlets after the experiment ended by liking, tweeting, or retweeting content that mentions a news source. When scraping user accounts after the experiment, we collect measures of engagement including likes, tweets, and retweets. We define *news engagement* and *non-news engagement* as interactions where the user engages with content related to a news outlet or not. Specifically, we classify a tweet as news engagement if it is in reply to a news outlet or mentions the news outlet directly. We classify likes and retweets as news engagement if they entail liking/retweeting a tweet that is classified as engaging with the publisher or liking/retweeting a tweet by the news outlet.

Figure 7 plots the effect of the disclosure treatment on news engagement (Panel A) and non-news engagement (Panel B). We find evidence that the disclosure treatment increases engagement with news outlets: retweets, likes, and tweets differentially increase in the treatment group. However, we find no evidence that the disclosure treatment increases engagement with non-news publishers.

To investigate how the disclosure treatment influences the composition of publishers with which users engage *after* the experiment, we calculate the treatment effect on total engagement with news publishers based on whether the participant began following the publisher during the experiment, whether the news outlet was suggested to the user during the experiment, and whether the user followed the publisher at baseline. The estimates plotted in Figure 8 indicate that users in the disclosure treatment are more likely to continue engaging with publishers they followed before and during the experiment after it ended.

The findings presented in this section suggest the disclosure treatment has important effects on long-term engagement. Participants in the disclosure treatment group are not simply following or unfollowing publishers to curate a news diet summary and then immediately reverting to their original news diet. The changes they make persist for months after the experiment, and they engage more with the publishers they follow. Together, this suggests there are important frictions in the formation of Twitter networks that could be driven by switching or search costs. Future studies should determine why these effects persist, given that the incentives users face after the experiment are identical.

5 The Role of Peer Information

In this section, we study the extent to which information that updates a participant's beliefs about the news diet of their network causes them to modify their news diets. We do so by exploiting variation across the treatment and control groups in: the peer information condition (see Section 2.1), the accuracy of the participants' prior beliefs about the news diets of their network, and the accuracy of the feedback provided to them during the experiment (see Section 2.4). We study whether this variation induces changes in posterior beliefs about the news diets of participants' networks and subsequent impacts on which news outlet a user follows.

5.1 Are News Choices Linked to Beliefs About Peers' News Diets?

We compare the treatment and control groups in the peer information condition (see Section 2.1). The only difference between these two groups is that the former receives feedback on the slant of their average follower's news diet. We estimate equation (1) to test whether receiving this information makes participants more likely to update their beliefs about their followers and more likely to modify their own news diets.

Figure 9 illustrates the results of this exercise. The upper left panel displays the fraction of people who updated their posterior beliefs about their followers' news diets across the treatment and control groups. We find that even though the control group received no explicit information about their network during the experiment, around 37% of the participants in the control group updated their beliefs. Yet participants in the treatment group are 32.3% (standard error of 6.2%) more likely than those in the control group to update their beliefs.

The remaining three panels in the figure visualize changes in users' news diets according to which peer treatment group they were assigned to (as in Figure 3). The upper right panel shows the probability of making at least one change to the news diet. The lower left panel indicates the change in the total number of outlets followed. The lower right panel displays the absolute value in the change of the slant of the followed outlets relative to the outlets participants were following before the randomization. We find no evidence that the peer treatment significantly impacts participants' news diets. We observe that the peer treatment group is less likely to make changes to their news diets, though this difference is not statistically distinguishable from zero. These ITT estimates are small in magnitude and not statistically distinguishable from zero. For instance, participants in the peer treatment group are 6.6% less likely than those in the control group to make at least one change to their news diet. This is a smaller magnitude compared to the 32.3% effect on the probability of updating posterior beliefs and the 21.4% effect of the disclosure condition treatment on the same outcome (discussed in Section 4.2).

The results thus far suggest that participants who receive feedback in the peer treatment group are more likely to update their beliefs, but this does not induce users to change their news diets. However, it is possible that people change their news consumption patterns in response to the new information. This feedback can update people's posteriors in different directions (depending on initial priors), so people could in theory make heterogeneous news consumption choices that cancel each other out perfectly when averaged across all participants. We exploit variations in the direction and magnitude of the update to investigate this issue.

To focus on "directional" outcomes (updates toward the left or right), we distinguish between feedback that is to the right of the prior (right feedback) vs. to the left of the prior (left feedback). To make belief adjustments comparable between left and right feedback, we normalize by multiplying adjustments following left feedback by -1. In other words, we apply the following transformation:

$$\widetilde{y} = \begin{cases} y & \text{if feedback>prior} \\ -y & \text{if feedback<prior} \end{cases} \tag{3}$$

where $\widetilde{y}$ is the normalized variable. Table 2 presents our estimates of equation (1) on the "directional" outcomes of belief updates and changes in the slant of news diets. To make the estimates comparable across the table, we divide the outcomes by their standard deviation so that the β estimates represent the effect of the peer information treatment in standard deviations of the outcome. We estimate the treatment effect on whether users move in the direction of the feedback (Columns 3 and 4) and the magnitude of changes in beliefs and news diets (Columns 1 and 2).

Panel A reports the effect of the peer treatment on the normalized update in beliefs about the news diets of the participant's followers. We find that participants in the treatment group (those who observe information about their followers) are much more likely to update their beliefs about the news diets of their followers in the direction of the feedback. This effect is statistically significant and equivalent to 12.4% (standard error of 1.8%) of a standard deviation of the outcome. In Column 2, we interact the treatment with the magnitude of the update (the absolute value of the difference between the feedback and the participant's self-reported prior), which we demean to facilitate interpretation. We find that when the distance between the prior and the feedback increases by one standard deviation (so that people's beliefs are less accurate), the treatment effects more than double (from 12.6% to 32.6%). Our estimates also imply that the treatment effect is negligible when individuals hold accurate beliefs about their peers (coefficient of 0.012 standard error of 0.031). Moreover, the results are consistent when studying the direction of belief updating in Columns 3 and 4. These results suggest that participants consistently update their beliefs about their peers in the direction of the feedback.

Panel B reports the treatment effect on the normalized change in the average slant of the participant's news diet. We find no evidence that participants in the treatment group make different changes to the slant of publishers they follow relative to the control group. In Column 3, the effect is -0.8% (standard error of 1.7%) of a standard deviation of the distribution (of the intensive margin) of slant change. This effect is not statistically distinguishable from zero and is much smaller in magnitude than the effect on beliefs in Panel A. Moreover, Column 2 shows that, unlike with beliefs, the accuracy of the prior does not explain changes in participants' news diets between the treatment and control groups. The results are consistent when looking at the magnitude of changes in the slant of participants' news diets (Columns 1 and 2).

These findings are similar and robust when exploiting a different source of variation to estimate the effect of changes in beliefs about peers' news diets – the accuracy of the feedback. As discussed in Section 2, the feedback provided to participants is an unbiased but noisy estimate of the true average slant of their followers' news diets; it is calculated as the average slant of five randomly sampled followers' news diets. This methodology has the advantage of generating exogenous variation in feedback, conditional on the true average slant of the news diet of the participant's peers. This implies that we do not need to rely only on the variation from the peer information randomization to identify the causal effects of changes in individual beliefs on subsequent consumption choices. We exploit this exogenous variation in feedback by estimating equation (2) separately for the control and treatment groups of the peer information condition. We interpret the results from the control group as a placebo, as this relies on information about the feedback they would have received had they been assigned to the treatment group.

Table 3 presents the results for the sample of people assigned to the peer treatment (Panels A and B) and peer control groups (Panels C and D). We focus on the same set of outcomes as in Table 2 in its unnormalized version. The coefficient γ measures the extent to which people update their beliefs about the news diets of their peers and their own news diets to the right (left) when receiving noisy information that the followers' ideology is to the right (left) of the true location.The coefficient in Panel A, Column 1 implies that participants who receive a positive (negative) sampling error in their feedback (i.e. the noisy average shown to participants is to the right of the more precise estimate of their peers' news diet) update their posterior beliefs about the news diets of their peers positively by 24.4% (standard error of 2.2%) of a standard deviation to the right (left) . Column 2 indicates that this effect is bigger when the noise is larger. This estimate also implies that the effect dissipates when participants receive an accurate signal (coefficient of 0.041 standard error of 0.029). Columns 3 and 4 present consistent results for the intensive margin. Panel B illustrates the treatment effect on news consumption changes. The coefficient in Panel B, Column 1 implies that individuals who receive information that their peers' ideology is to the right of the more precise estimate choose news outlets that have a slant of 0.4% (standard error of 1.9%) of a standard deviation to the left. This coefficient is statistically indistinguishable from zero and small compared to the magnitude of beliefs. Nor do we find evidence that this effect varies based on the intensity of the signal (Column 2 in Panel B). Columns 3 and 4 in Panel B present similar findings on the direction of changes in the slant of the participant's news diet.

When estimating these equations for the placebo sample (participants who did not receive the feedback, Panels C and D), we find that, despite the smaller sample, nearly all the coefficients are statistically indistinguishable from zero and smaller in magnitude compared to those in Panel A.[15] This is true for both changes in belief and slant.[16]

Despite utilizing variation from a different source, the results remain remarkably con-

[15]One of the 12 estimates in Panels C and D (the interaction term in Column 4 of Panel C) is statistically significant. This 8.3% rejection rate is consistent with what we would expect from a two-sided test on the basis of sampling variation.

[16]Appendix Table A-2 reports the results when we combine variation on the random variation of the peer information condition and the feedback (rather than the accuracy of the prior) following the same approach as in Table 2 for comparability. As expected, the results are very consistent.

sistent across both methodologies. These analyses present a clear insight: individuals significantly adjust their beliefs after receiving information about the slant of their followers' news diets. However, despite this notable shift in beliefs, we find limited evidence of a significant change in people's news choices in response to this update.

5.2 Followers vs. Followings

One potential explanation for the findings presented above is that people update their news diets as a function of the ideology of the news diets of who they follow (their followings) rather than who follows them (their followers). Such asymmetric behavior could be due to users having to make an active choice to follow someone; they have less control over who follows them on the platform. This difference is unlikely to explain why we find that participants' beliefs about the news diets of their peers has little or no effect on their own news diets. Platforms like Twitter typically have a large overlap between who follows a user and who the participant follows. As a result, the information about the slant of the news diet of the participant's followers should be highly correlated with the slant of the news diet of the users the participant follows. In this section, we investigate this relationship more formally.

Our sample contains a high share of mutual connections on average, and high variation in this ratio across participants. We define the share of mutual connections as the ratio between the number of mutual connections and the maximum between the number of followers and followings. Let E_j be the set of accounts that follow j (the followers) and I_j the set of accounts that j follows (the followings). The mutual connections share of participant j is then defined as:

$$\text{Mutual Connections Share}_j = \frac{\sum_{e \in E_j} \sum_{i \in I_j} \mathbb{I}[i = e]}{\max(|E_j|, |I_j|)} \tag{4}$$

The average participant has a mutual connection share of 44.3% and the median participant has a mutual connection share of 44.2%. This quantity ranges from 8.2% for the 10th percentile and 89.9% for the 90th percentile.

To test whether beliefs about the followings, rather than the followers, help explain news consumption choices, we compare the treatment effects across individuals with a low vs. high mutual connections share. If individuals are aware of these mutual connections, we should expect the latter group to be more likely to update their beliefs about their followings. Descriptive evidence suggests that participants are indeed aware of these connections. For instance, we detect a very strong negative correlation between the mutual connections share and the distance between the participant's prior beliefs about the ideology of their followers and followings: compared to individuals with no mutual connections, participants whose followers and followings are identical report, on average, a reduction in this distance equivalent to 44.9% of a standard deviation.

We study this issue more rigorously by estimating the effect of the treatment in the peer information condition (which provides participants with information about their followers) on belief updating about the slant of the news diets of the accounts they follow. We also estimate an extended version of this equation to consider heterogeneous treatment effects by the mutual connection share. Rational individuals who update their beliefs about the slant of their followers' news diets should also update their beliefs about the ideology of the accounts they follow when they receive a signal about the ideology of their followers. This should be especially pronounced for participants with a high mutual connection share.

Table 4 reports the main results using variation of the accuracy of the prior beliefs (as in Table 4). Panel A reports the treatment effect on normalized beliefs adjustment. The estimate in Column 1 implies that participants do in fact update their beliefs about their following when they are confronted with information about their followers. This effect is statistically significant and equivalent to 4.6% (s.d. 0.019) of a standard deviation of the outcome. Column 2 indicates that participants are significantly more likely to update their beliefs about their followings if there is a large overlap between their followers and followings. Yet our estimates in Column 2 imply that individuals do not update their beliefs about their followings (the coefficient is 0% standard error of 2.6%) if the network connectedness is zero. This suggests that individuals do *not* use information about their followers to make inferences about their followings if there is no overlap between the two groups. The fact that the point estimates in these two columns (and in Columns 3 and 4 when looking at the extensive margin) are smaller in magnitude than those reported in Table 2 is also reassuring, as the signal should always be more informative about the followers than the followings (except, of course if network connectedness is one).

Using a similar strategy, in Panel B we study whether changes in beliefs about the ideology of one's followings lead to changes in consumption choices by looking at the effect on the normalized slant change. The main coefficient of interest is the interaction between the treatment and the network connectedness reported in the even columns because, as we show above, this interaction identifies participants who considerably update their beliefs about the followings. The estimates in Columns 2 and 4 are small in magnitude and not statistically significant, which suggests the changes in beliefs about the followings are unlikely to affect news consumption choices.[17]

5.3 Persistence of Treatment Effects and Impact on Engagement

We analyze the persistence of the peer information treatment and its effect on engagement as in Section 4.4. Figure A-10 plots the effect of the peer treatment over time and the null effect persists. In addition, we find no evidence that the peer treatment impacts engagement with either news publishers or non-news publishers (Figure A-11). Our findings therefore reject the hypothesis that users' news diets only incorporate information about their peers over time, as we find no change in the publishers users choose to follow or engage with after the experiment.

6 Conclusion

There is a widespread belief that interactions with like-minded peers tend to limit exposure to differing viewpoints and can reinforce existing beliefs, thus polarizing societies. However, there is limited empirical evidence of how these interactions lead to changes in the demand for biased information and, ultimately, beliefs. Even less is known about the mechanisms through which this happens. In this paper, we study the role of two important mechanisms – social image concerns and peer effects – in explaining preferences for biased news. Our field experiment on Twitter induced variation in both an individual's perceptions of the political leanings of their peers' news consumption and the visibility of their own news preferences to their social media followers. We find that individuals significantly adjust their news consumption choices when they believe they are being observed by their peers, which suggests the social image concern mechanism is an important channel. However, we find little support for the peer effects channel: changes in beliefs about the political ideology of the news that peers consume has little impact on individuals' own news consumption choices.

Our study has important policy implications. It provides evidence that interactions with peers (potentially likeminded peers within an echo chamber) can increase the demand

[17]Appendix Table A-3 establishes that we obtain very similar conclusions when using the additional approach that exploits exogenous variation in the accuracy of the feedback.

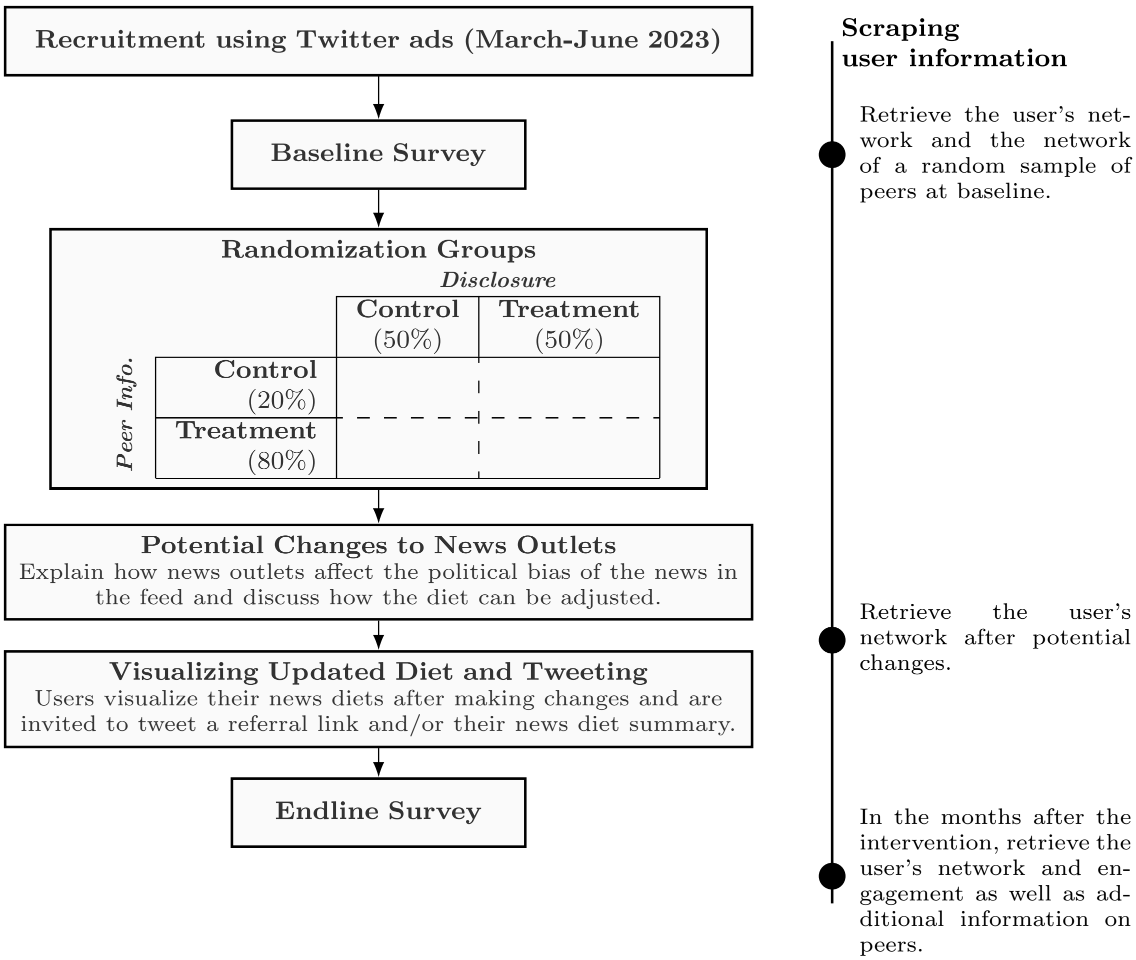
for more moderate content. As we demonstrate, individuals care about how their peers perceive the content they engage with. Furthermore, users significantly value demonstrating to their peers that they consume more moderate news relative to their private bliss point . Therefore, by amplifying the visibility of user interactions, social media can help moderate the content that users choose to consume. Importantly, our results indicate that these effects are not only statistically significant but also substantial in magnitude. Thus, they must be carefully considered in the formulation of policies designed to mitigate polarization. To the extent that encouraging users to consume balanced news diets is a desirable goal, our results suggest clear policies that would encourage this behavior. Given the significant influence of the social image concern channel and the moderating impact of the incentive to publicize news diets, fostering enhanced transparency on platforms could yield favorable effects on users' news consumption patterns. Subtle tools that weigh user preferences for privacy and foster transparency of news choices on social media can have net desirable effects.[18]

Despite the evidence that social media mitigates the demand for polarized news content in our setting, further research is needed to understand social media's aggregate effect on polarization. On the one hand, social media is not only believed to affect polarization through the creation of echo chambers. To maximize engagement, algorithms in these technologies are believed to prioritize users' previous behavior, thus limiting their exposure to counter-attitudinal content. The degree to which social media algorithms expose users to segregated content, and how this exposure contributes to polarization, form a central pillar of debate in recent research. Further evidence is needed to deepen our understanding of this phenomenon [González-Bailón et al., 2023, Guess et al., 2023]. On the other hand, there may be other channels through which interactions with others on social media platforms could increase polarization. For example, peer effects could change participants' behavior and be uncorrelated with the variation in peers' ideological positions or other forms of social image concerns (e.g. peers may expect individuals to advocate their support for radical issues), orthogonal to information on the ideology of news choic es. The salience of these potential mechanisms is also an important direction for future research.

[18] The blindspotter tool by ground.news is one such example `https://ground.news/blindspotter/twitter`.

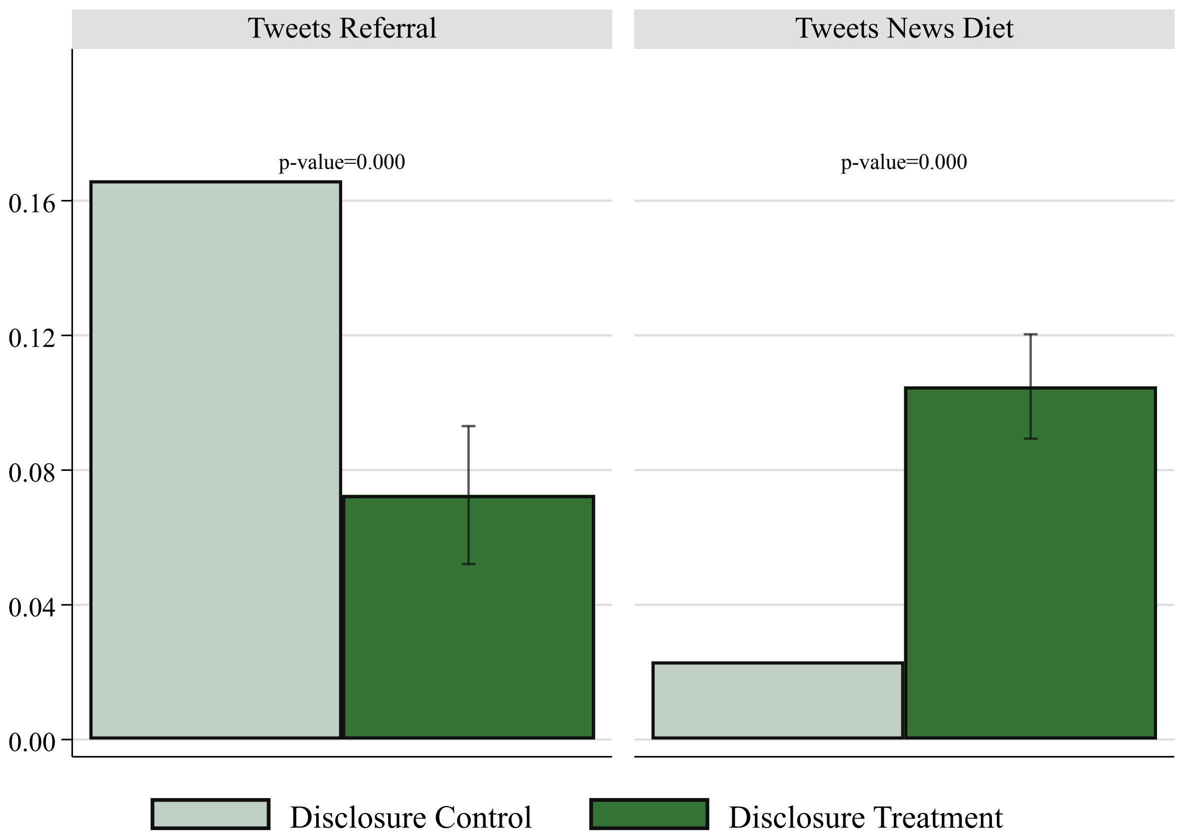
7 Figures and Tables

Figure 1: Experimental Design



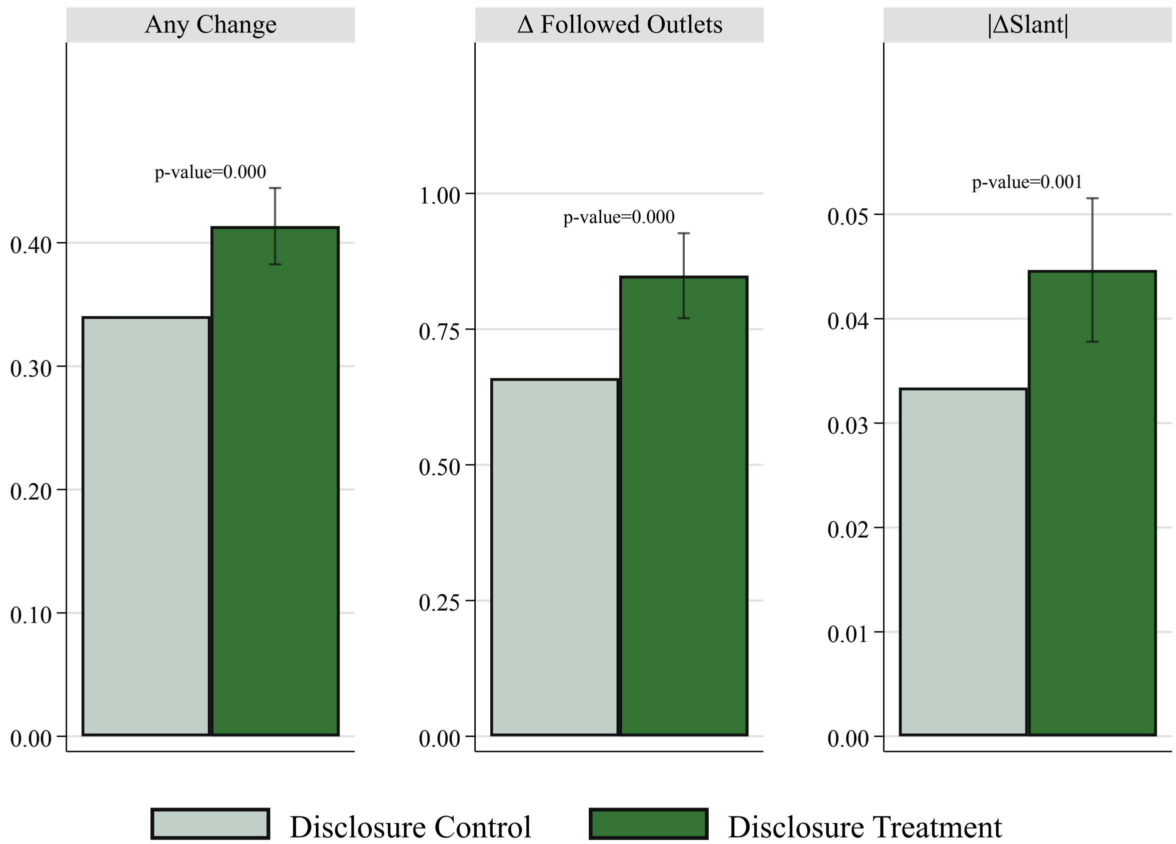
Notes: This figure shows the experimental design of the study. The left side shows the process that participants follow to complete the experiment and the randomization assignments. The right side highlights the periods in which we scrape information about the participants and/or their peers about accounts followed and/or engagement. A total of 954,913 unique Twitter users were shown the recruitment ads. 12,940 clicked on the ads, 5,192 consented to participate in the study and 3,757 finished the study.

Figure 2: Disclosure Condition and Compliance Rates



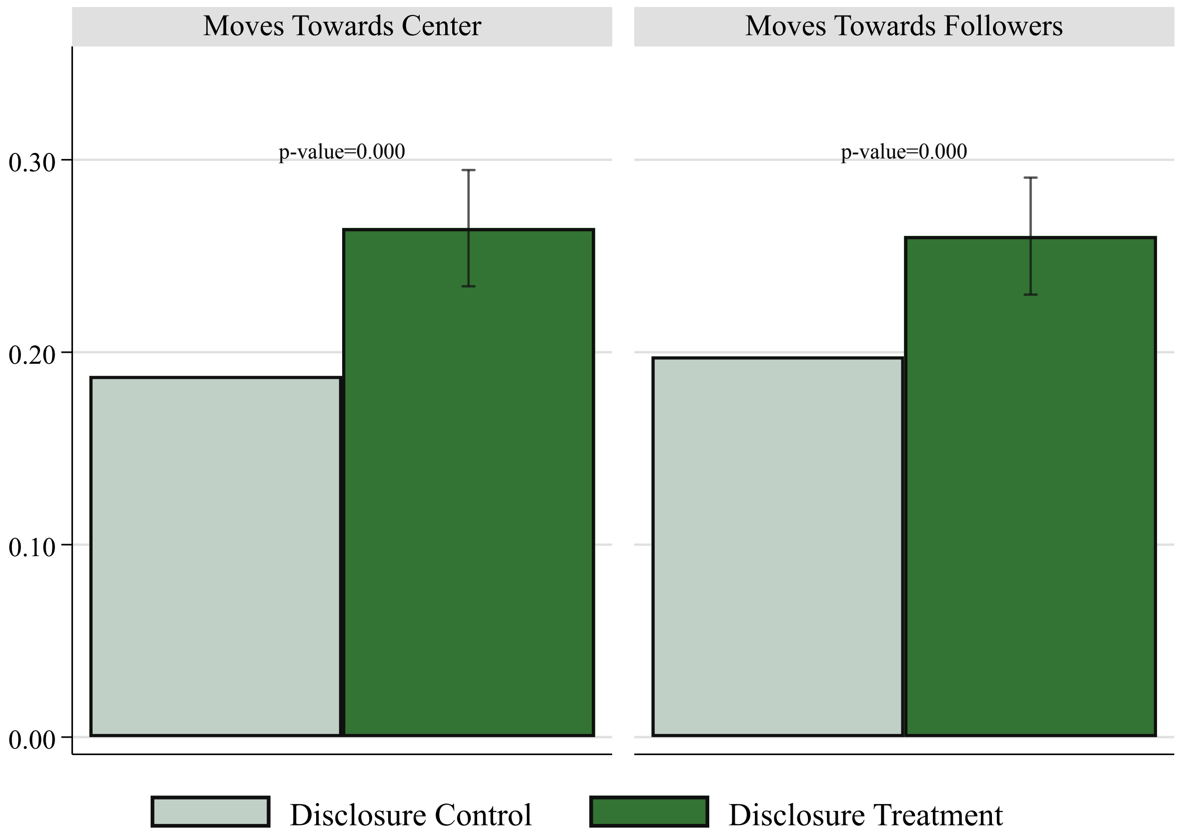
Notes: This figure reports the fraction of individuals that tweet either a referral link with no personal information (left panel) or the news diet summary (right panel) across the treatment and control groups in the disclosure condition. The difference between the treatment and control groups in the disclosure condition is that only the former is asked to reveal information about which news outlets they follow to their peers, which is done after they have the chance to modify these outlets (see more details in Section 2.1). The whiskers indicate the 95% confidence intervals of the difference between the effect in the treatment and control groups following equation (1). We also report the p-value of the null hypothesis of no difference between the treatment and control groups.

Figure 3: Disclosure Condition and Changes to News Outlets Followed



Notes: This figure reports the mean of the following three outcomes across participants assigned to the treatment and control groups in the disclosure condition: an indicator variable if the participant makes any change to the news outlets they follow (left), the change in the number of news outlets followed (center), and the change in the absolute slant of the news outlets followed (right). In all cases, we compare these outcomes in the post-intervention period relative to baseline. The difference between the treatment and control groups in the disclosure condition is that only the former is asked to reveal information about which news outlets they follow to their peers, which is done after they have the chance to modify these outlets (see more details in Section 2.1). The whiskers indicate the 95% confidence intervals of the difference between the effect in the treatment and control groups following equation (1). We also report the p-value of the null hypothesis of no difference between the treatment and control groups.

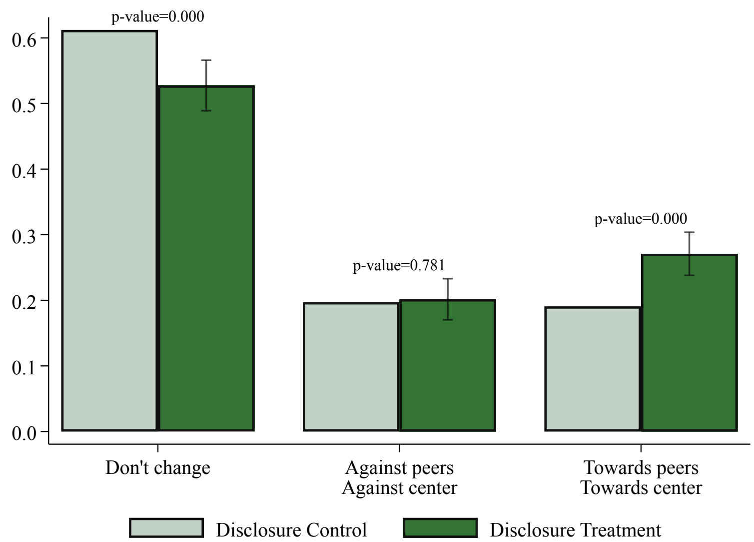
Figure 4: Disclosure Condition and Movements Toward the Center vs. Peers



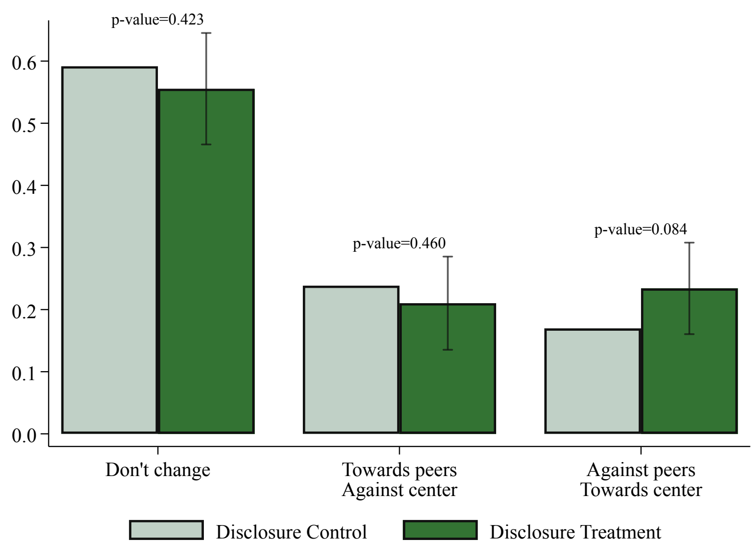
Notes: This figure reports the fraction of individuals that adjust the slant of their news outlets followed toward the center (left) vs. toward their peers (right) across the treatment and control groups in the disclosure condition. The difference between the treatment and control groups in the disclosure condition is that only the former is asked to reveal information about which news outlets they follow to their peers, which is done after they have the chance to modify these outlets (see more details in Section 2.1). The whiskers indicate the 95% confidence intervals of the difference between the effect in the treatment and control groups following equation (1). We also report the p-value of the null hypothesis of no difference between the treatment and control groups.

Figure 5: Disclosure Condition and Movements Toward the Center vs. Peers – News Consumers

A. Peers and Center in the Same Direction

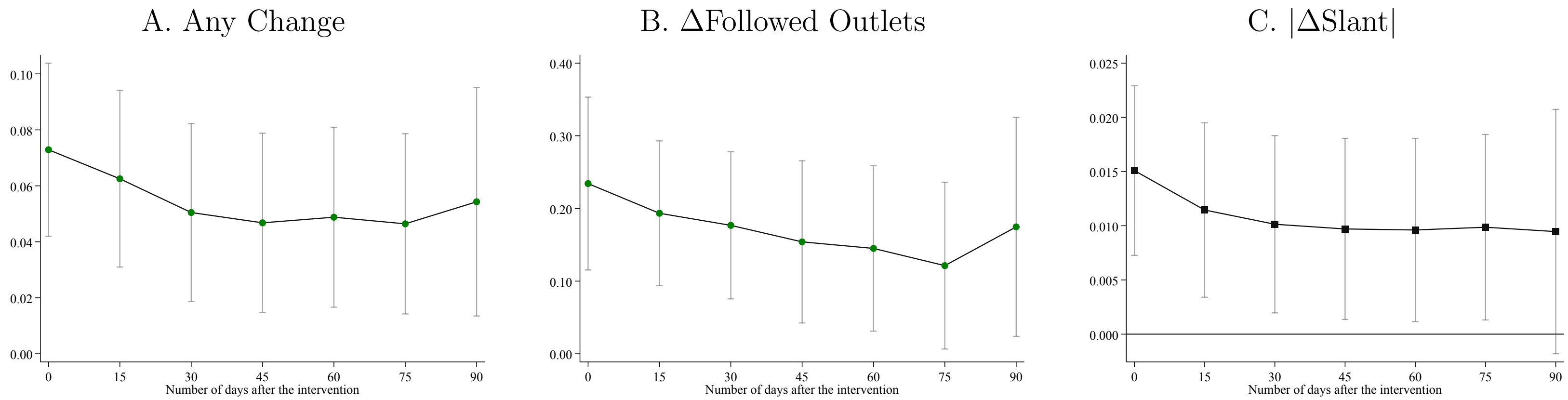


B. Peers and Center in Opposite Directions



Notes: This figure reports the treatment effects of the disclosure condition on whether participants adjust the slant of their news outlets followed toward the center vs. their peers for news consumers (participants who follow at least one news outlet, 80% of our sample). Panel A displays the results for participants whose peers and the center are in the same direction: Democrats (Republicans) whose peers are to their right (left). Panel B presents the results for participants whose peers and the center are in opposite directions: Democrats (Republicans) whose peers are to their left (right). The difference between the treatment and control groups in the disclosure condition is that only the former is asked to reveal information about which news outlets they follow to their peers, which is done after they have the chance to modify these outlets (see more details in Section 2.1). The whiskers indicate the 95% confidence intervals of the difference between the effect in the treatment and control groups following equation (1). We also report the p-value of the null hypothesis of no difference between the treatment and control groups.

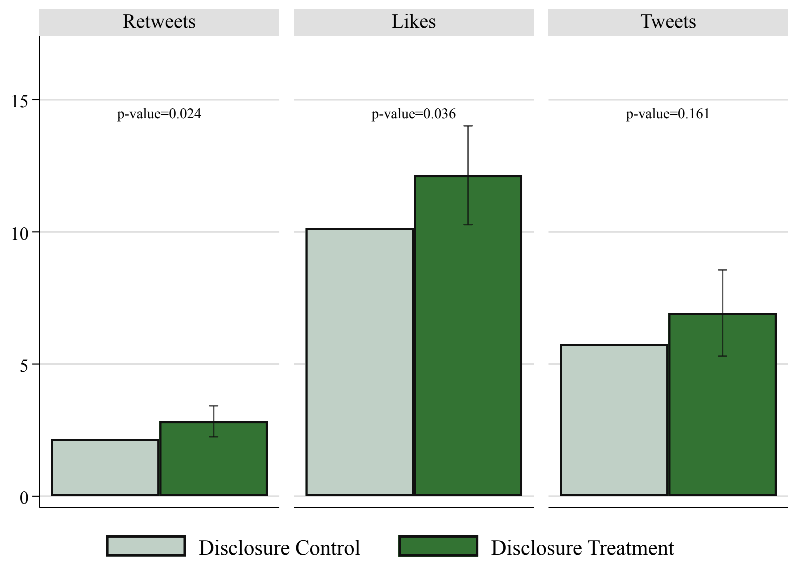
Figure 6: Disclosure Condition and Treatment Effect Over Time



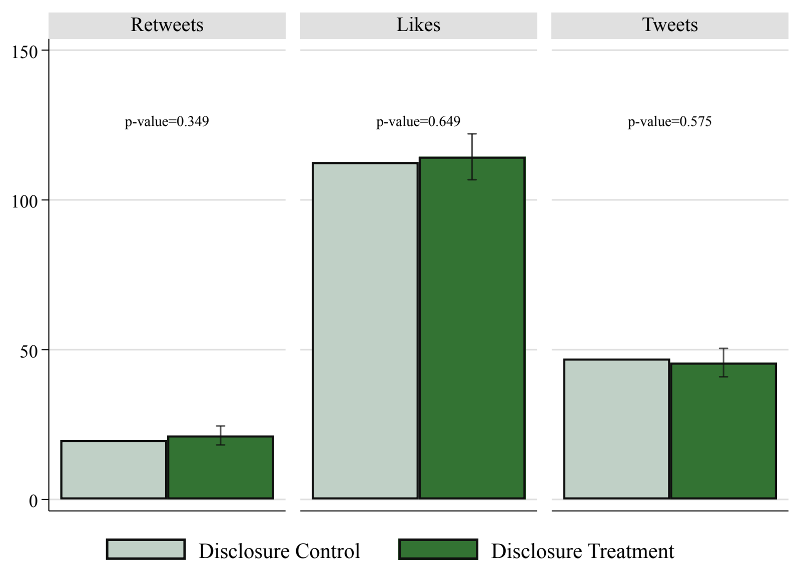
Notes: This figure reports ITT estimates (comparing the treatment and control groups in the disclosure condition) on: an indicator variable if the participant makes any change to the news outlets they follow (left), the change in the number of news outlets followed (center), and the change in the absolute slant of the news outlets followed (right). In all cases, we compare these outcomes at different points in time in the post-intervention period (see x-axis) vs. baseline. The difference between the treatment and control groups in the disclosure condition is that only the former is asked to reveal information about which news outlets they follow to their peers, which is done after they have the chance to modify these outlets (see more details in Section 2.1). The whiskers indicate the 95% confidence intervals of the difference between the effect in the treatment and control groups following equation (1).

Figure 7: Disclosure Condition and Engagement

A. News Engagement

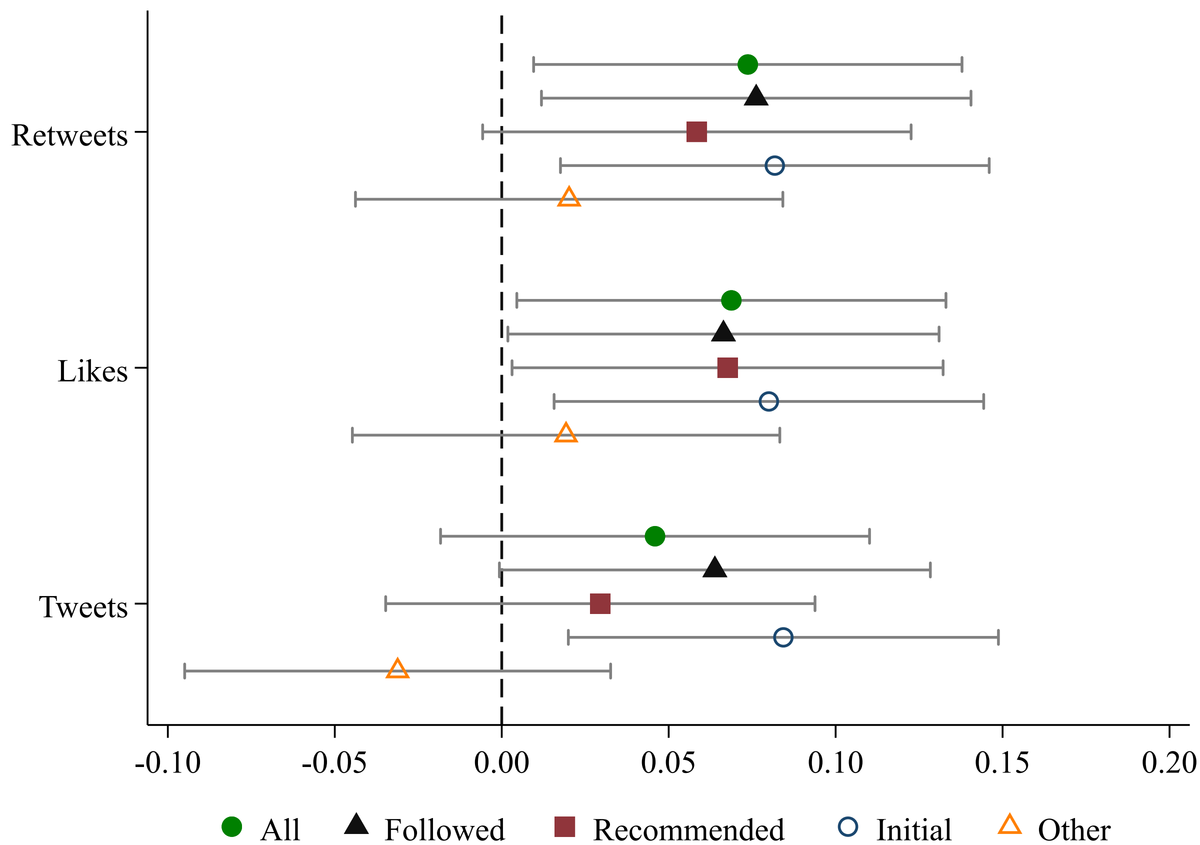


B. Non-News Engagement



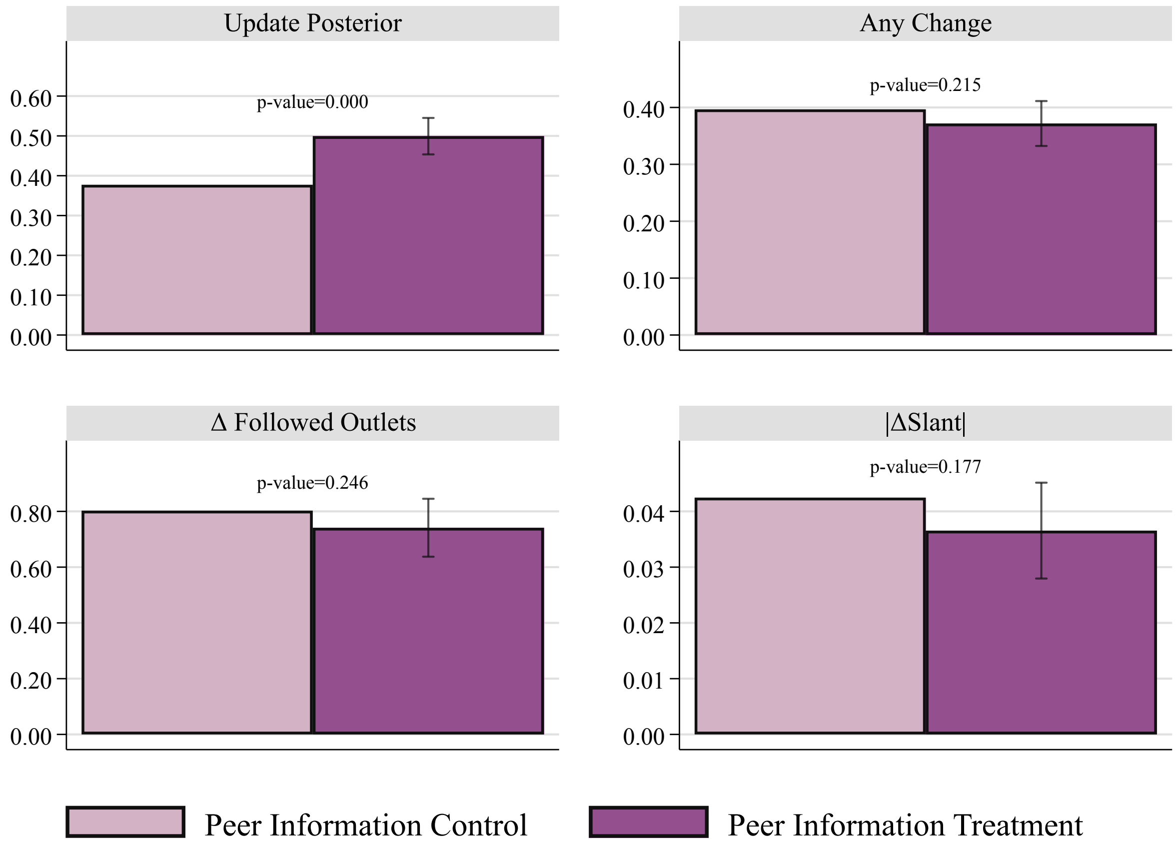
Notes: This figure reports the mean of the number of retweets, likes, and tweets associated with news outlets (Panel A) and non-news outlets (Panel B) across participants assigned to the treatment and control groups in the disclosure condition. Retweets, likes, and tweets are defined as associated with news outlets if the tweet mentions a news outlet or responds to a tweet created by a news outlet. The difference between the treatment and control groups in the disclosure condition is that only the former is asked to reveal information about which news outlets they follow to their peers, which is done after they have the chance to modify these outlets (see more details in Section 2.1). The whiskers indicate the 95% confidence intervals of the difference between the effect in the treatment and control groups following equation (1). We also report the p-value of the null hypothesis of no difference between the treatment and control groups.

Figure 8: Disclosure Condition and News Engagement with Different Types of News Outlets



Notes: This figure reports ITT estimates (comparing the treatment and control groups in the disclosure condition) of the number of retweets, likes, and tweets associated with the following subsets of news outlets: Any (All), those followed during the intervention (Followed), the six recommended publishers (Recommended), the set of news outlets at baseline (Initial), and the set of news outlets neither initially followed nor recommended (Other). The outcomes are standardized, so the plotted coefficients represent ITT estimates in standard deviation units of the outcome. Retweets, likes, and tweets are defined as associated with news outlets if the tweet mentions a news outlet or responds to a tweet created by a news outlet. The difference between the treatment and control groups in the disclosure condition is that only the former is asked to reveal information about which news outlets they follow to their peers, which is done after they have the chance to modify these outlets (see more details in Section 2.1). The whiskers indicate the 95% confidence intervals of the difference between the effect in the treatment and control groups following equation (1).

Figure 9: Peer Information Condition, Belief Updating, and News Outlets Followed



Notes: This figure reports the mean of four outcomes across participants assigned to the treatment and control groups in the peer condition: an indicator variable if participants update their beliefs regarding the slant of the news consumed by their peers (upper left), an indicator variable if participants make any changes to the news outlets followed (upper right), the change in the number of news outlets followed (bottom left), and the change in the absolute slant of the news outlets followed (bottom right). In all cases, we compare these outcomes in the post-intervention period vs. baseline. The difference between the treatment and control groups in the peer information condition is that only the former receives feedback on the slant of a random sample of their peers/followers (more details are available in Section 2.1). The whiskers indicate the 95% confidence intervals of the difference between the effect in the treatment and control groups following equation (1). We also report the p-value of the null hypothesis of no difference between the treatment and control groups.

Table 1: Descriptive Statistics

	(1)	(2)
	Sample	US adults
Male	0.68	0.49
Age	50.49	47.6
White	0.91	0.79
Graduate degree	0.49	0.14

Notes: Columns 1 and 2 report average demographics for our final sample and for the US adult population, respectively.

Table 2: Peer Information Condition and Belief Adjustment on Followers and Subsequent Changes to News Choices – Variation on Accuracy of Prior Beliefs

	(1)	(2)	(3)	(4)
	A. Normalized Belief Adjustment – Followers			
	Belief Change		Sign(Belief Change)	
Peer Treatment	0.124	0.126	0.140	0.142
	(0.018)	(0.018)	(0.019)	(0.019)
Peer Treatment × Distance Prior vs Feedback		0.200		0.172
		(0.055)		(0.040)
	B. Normalized Adjustment in Slant of Outlets			
	Slant Change		Sign(Slant Change)	
Peer Treatment	-0.011	-0.011	-0.008	-0.008
	(0.019)	(0.019)	(0.017)	(0.017)
Peer Treatment × Distance Prior vs Feedback		0.023		-0.016
		(0.043)		(0.037)
Observations	3,757	3,757	3,757	3,757

Notes: This table reports ITT estimates (comparing the treatment and control groups in the peer condition) on normalized belief adjustment about the *followers* (Panel A) and normalized adjustments in the slant of news outlets (Panel B) following equation (1). Belief or slant adjustments are defined as the difference between the post-intervention and baseline period. To make belief adjustments comparable between left and right feedback, we normalize by multiplying adjustments following left feedback by -1 (see equation 3). For each set of outcomes, we report both the intensive (how much participants move in the direction of the feedback, Columns 1 and 2) and extensive margin (whether participants move in the direction of the feedback, Columns 3 and 4). The even columns present heterogeneous treatment effects by the absolute value of the difference between the prior and the feedback. The outcomes as well as the distance between the prior and the feedback are standardized. The difference between the treatment and control groups in the peer information condition is that only the former receives feedback on the slant of a random sample of their peers/followers (more details are available in Section 2.1). Standard errors reported in parentheses are robust against heteroscedasticity.

Table 3: Peer Information Condition and Belief Adjustment on Followers and Subsequent Changes to News Choices – Exogenous Variation on the Accuracy of the Feedback

	(1)	(2)	(3)	(4)
A. Peer Treatment Group Sample, Beliefs	Belief Adjustment – Followers			
	Belief Change		Sign(Belief Change)	
Feedback > Truth	0.244	0.254	0.224	0.234
	(0.022)	(0.021)	(0.021)	(0.021)
(Feedback > Truth) × Distance Truth vs Feedback		0.302		0.279
		(0.033)		(0.029)
B. Peer Treatment Group Sample, News Choices	Normalized Adjustment in Slant of Outlets			
	Slant Change		Sign(Slant Change)	
Feedback > Truth	-0.004	-0.004	-0.006	-0.004
	(0.019)	(0.019)	(0.019)	(0.019)
(Feedback > Truth) × Distance Truth vs Feedback		0.024		0.009
		(0.026)		(0.028)
Observations	2,953	2,953	2,953	2,953
C. Peer Control Group Sample, Beliefs	Belief Adjustment – Followers			
	Belief Change		Sign(Belief Change)	
Feedback > Truth	0.021	0.020	0.046	0.045
	(0.041)	(0.041)	(0.042)	(0.042)
(Feedback > Truth) × Distance Truth vs Feedback		0.090		0.132
		(0.063)		(0.059)
D. Peer Control Group Sample, News Choices	Normalized Adjustment in Slant of Outlets			
	Slant Change		Sign(Slant Change)	
Feedback > Truth	0.008	0.008	0.034	0.034
	(0.037)	(0.037)	(0.038)	(0.038)
(Feedback > Truth) × Distance Truth vs Feedback		-0.028		-0.072
		(0.031)		(0.050)
Observations	722	722	722	722

Notes: This table reports estimates from equation (2) on beliefs about followers (Panels A and C) and slant adjustments (Panels B and D), defined as the difference between the post-intervention and the baseline period. Panels A and B (C and D) report the results for the sample of participants assigned to the treatment (control) group in the peer condition. $Feedback > Truth$ is an indicator variable equal to 1 if the slant of news outlets followed by a participant's peers is greater (to the right) when constructed based on a random sample of five peers (the feedback) compared to when constructed based on a larger sample of peers (the truth, see Section 2.4). For each set of outcomes, we report both the intensive (how much participants move to the right, Columns 1 and 2) and extensive margin (whether participants move to the right, Columns 3 and 4). The even columns present heterogeneous treatment effects by the absolute value of the difference between the feedback and the truth. The outcomes as well as the distance between the feedback and the truth are standardized. The difference between the treatment and control groups in the peer information condition is that only the former receives feedback on the slant of a random sample of their peers/followers (more details are available in Section 2.1). Standard errors reported in parentheses are robust against heteroscedasticity.

Table 4: Peer Information Condition and Belief Adjustment on Followings and Subsequent Changes to News Choices – Variation in Accuracy of Prior Beliefs

	(1)	(2)	(3)	(4)
	A. Normalized Belief Adjustment – Followings			
	Belief Change		Sign(Belief Change)	
Peer Treatment	0.046	-0.000	0.064	0.028
	(0.019)	(0.026)	(0.019)	(0.026)
Peer Treament × Mutual Connections Share		0.108		0.085
		(0.045)		(0.040)
	B. Normalized Adjustment in Slant of Outlets			
	Slant Change		Sign(Slant Change)	
Peer Treatment	-0.011	0.006	-0.008	0.008
	(0.019)	(0.026)	(0.017)	(0.023)
Peer Treament × Mutual Connections Share		-0.039		-0.037
		(0.037)		(0.035)
Observations	3,757	3,757	3,757	3,757

Notes: This table reports ITT estimates (comparing the treatment and control groups in the peer condition) on normalized belief adjustment about the *followings* (Panel A) and normalized adjustments in the slant of news outlets (Panel B) following equation (1). Belief or slant adjustments are defined as the difference between the post-intervention and the baseline period. To make belief adjustments comparable between left and right feedback, we normalize by multiplying adjustments following left feedback by -1 (see equation 3). For each set of outcomes, we report both the intensive (how much participants move in the direction of the feedback, Columns 1 and 2) and extensive margin (whether participants move in the direction of the feedback, Columns 3 and 4). The even columns present heterogeneous treatment effects by the share of mutual connections (fraction of accounts that are part of both the followers and followings sets, see equation (4)). The outcomes and the share of mutual connections are both standardized. The difference between the treatment and control groups in the peer information condition is that only the former receives feedback on the slant of a random sample of their peers/followers (more details are available in Section 2.1). Standard errors reported in parentheses are robust against heteroscedasticity.

References

H. Allcott, L. Braghieri, S. Eichmeyer, and M. Gentzkow. The welfare effects of social media. *American Economic Review*, 110(3):629–76, 2020.

H. Allcott, M. Gentzkow, and L. Song. Digital addiction. *American Economic Review*, 112 (7):2424–2463, 2022.

S. Aral and M. Zhao. Social media sharing and online news consumption. *Available at SSRN 3328864*, 2019.

S. Athey, M. Mobius, and J. Pal. The impact of aggregators on internet news consumption. Technical report, National Bureau of Economic Research, 2021.

E. Bakshy, S. Messing, and L. A. Adamic. Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348(6239):1130–1132, June 2015. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.aaa1160. URL `https://science-sciencemag-org.libproxy.mit.edu/content/348/6239/1130`. Publisher: American Association for the Advancement of Science Section: Report.

L. Boxell, M. Gentzkow, and J. M. Shapiro. Cross-country trends in affective polarization. *Review of Economics and Statistics*, pages 1–60, 2022.

D. Broockman and J. Kalla. The impacts of selective partisan media exposure: a field experiment with fox news viewers. *OSF Preprints*, 2022.

J. Bruner. Tweets loud and quiet. *O'Reilly Radar. Retrieved October*, 6:2015, 2013.

L. Bursztyn and R. Jensen. How does peer pressure affect educational investments? *The quarterly journal of economics*, 130(3):1329–1367, 2015.

L. Bursztyn and R. Jensen. Social image and economic behavior in the field: Identifying, understanding, and shaping social pressure. *Annual Review of Economics*, 9:131–153, 2017.

L. Bursztyn, T. Fujiwara, and A. Pallais. 'acting wife': Marriage market incentives and labor market investments. *American Economic Review*, 107(11):3288–3319, 2017.

L. Bursztyn, G. Egorov, and S. Fiorin. From extreme to mainstream: The erosion of social norms. *American economic review*, 110(11):3522–3548, 2020a.

L. Bursztyn, A. L. González, and D. Yanagizawa-Drott. Misperceived social norms: Women working outside the home in saudi arabia. *American economic review*, 110(10):2997–3029, 2020b.

A. Casas, E. Menchen-Trevino, and M. Wojcieszak. Exposure to extremely partisan news from the other political side shows scarce boomerang effects. *Political Behavior*, pages 1–40, 2022.

F. Chopra, I. Haaland, and C. Roth. Do people demand fact-checked news? evidence from us democrats. *Journal of Public Economics*, 205:104549, 2022.

F. Chopra, I. Haaland, and C. Roth. The demand for news: Accuracy concerns versus belief confirmation motives. *NHH Dept. of Economics Discussion Paper*, (01), 2023.

S. DellaVigna, J. A. List, U. Malmendier, and G. Rao. Voting to tell others. *The Review of Economic Studies*, 84(1):143–181, 2016.

R. Enikolopov, A. Makarin, and M. Petrova. Social media and protest participation: Evidence from russia. *Econometrica*, 88(4):1479–1514, 2020.

J. Fardouly and L. R. Vartanian. Social media and body image concerns: Current research and future directions. *Current opinion in psychology*, 9:1–5, 2016.

R. G. Fryer Jr and P. Torelli. An empirical analysis of 'acting white'. *Journal of Public Economics*, 94(5-6):380–396, 2010.

P. Funk. Social incentives and voter turnout: evidence from the swiss mail ballot system. *Journal of the European economic association*, 8(5):1077–1103, 2010.

M. Gentzkow and J. M. Shapiro. Media bias and reputation. *Journal of political Economy*, 114(2):280–316, 2006.

M. Gentzkow and J. M. Shapiro. What drives media slant? evidence from us daily newspapers. *Econometrica*, 78(1):35–71, 2010.

M. Gentzkow and J. M. Shapiro. Ideological segregation online and offline. *The Quarterly Journal of Economics*, 126(4):1799–1839, 2011.

A. S. Gerber, D. P. Green, and C. W. Larimer. Social pressure and voter turnout: Evidence from a large-scale field experiment. *American political Science review*, 102(1):33–48, 2008.

S. González-Bailón, D. Lazer, P. Barberá, M. Zhang, H. Allcott, T. Brown, A. Crespo-Tenorio, D. Freelon, M. Gentzkow, A. M. Guess, et al. Asymmetric ideological segregation in exposure to political news on facebook. *Science*, 381(6656):392–398, 2023.

A. Guess and A. Coppock. Does counter-attitudinal information cause backlash? results from three large survey experiments. *British Journal of Political Science*, 50(4):1497–1515, 2020.

A. M. Guess, N. Malhotra, J. Pan, P. Barberá, H. Allcott, T. Brown, A. Crespo-Tenorio, D. Dimmery, D. Freelon, M. Gentzkow, et al. How do social media feed algorithms affect attitudes and behavior in an election campaign? *Science*, 381(6656):398–404, 2023.

R. Levy. Social media, news consumption, and polarization: Evidence from a field experiment. *American economic review*, 111(3):831–70, 2021.

A. Lindbeck. Incentives and social norms in household behavior. *The American Economic Review*, 87(2):370–377, 1997.

S. Messing and S. J. Westwood. Selective exposure in the age of social media: Endorsements trump partisan source affiliation when selecting news online. *Communication research*, 41(8):1042–1063, 2014.

S. Mullainathan and A. Shleifer. The market for news. *American economic review*, 95(4):1031–1053, 2005.

K. Muralidharan, M. Romero, and K. Wüthrich. Factorial designs, model selection, and (incorrect) inference in randomized experiments. *The Review of Economics and Statistics*, pages 1–44, 2023.

B. Nyhan, J. Settle, E. Thorson, M. Wojcieszak, P. Barberá, A. Y. Chen, H. Allcott, T. Brown, A. Crespo-Tenorio, D. Dimmery, et al. Like-minded sources on facebook are prevalent but not polarizing. *Nature*, 620(7972):137–144, 2023.

R. E. Robertson, S. Jiang, K. Joseph, L. Friedland, D. Lazer, and C. Wilson. Auditing partisan audience bias within google search. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW):1–22, 2018.

J. E. Settle. *Frenemies: How social media polarizes America*. Cambridge University Press, 2018.

W. Suen. The self-perpetuation of biased beliefs. *The Economic Journal*, 114(495):377–396, 2004.

C. R. Sunstein. *# Republic: Divided democracy in the age of social media*. Princeton University Press, 2018.

D. J. Watts, D. M. Rothschild, and M. Mobius. Measuring the news and its impact on democracy. *Proceedings of the National Academy of Sciences*, 118(15):e1912443118, 2021.

A Appendix

A.1 Recruitment and Data Collection

We recruited participants between March and June 2023 using Twitter ads. The ads contained a map indicating whether the majority of votes in the 2020 presidential election in every US county was obtained by the Republican Party (red) or Democratic Party (blue), and text that varies whether a monetary or non-monetary incentive (participants learning about their own political bias) is offered (see Appendix Figure A-12). We find that participants strongly responded to the message inviting them to learn about their political bias. The ad in Panel A (no monetary incentive) has the lowest acquisition cost per click ($0.13, compared to our average acquisition of $0.15) and the highest link-click-rate (1.47%, compared to the average of 1.36%).

A.2 Construction of Random Sample of Twitter Users

It is often useful to compare Twitter users in our data to a random sample of the population, though constructing a random sample of Twitter users is not trivial. We therefore use the data provided alongside Bruner [2013] to construct a random sample of Twitter users. Here, we briefly describe how Bruner [2013] constructs their random sample. Twitter assigns each account a unique integer ID. Until 2015, this ID was approximately sequential, so at the time it was feasible to achieve a random sample of Twitter accounts by randomly sampling Twitter IDs from the known range of possible IDs. After 2015, Twitter changed its ID assignment algorithm, making the ID space substantially more sparse and this was no longer computationally feasible. Therefore, we use a random sample of Twitter accounts from 2013 that was created by randomly querying Twitter IDs in the ID space. This sample of 400,000 accounts was published alongside Bruner [2013]. Given API limitations, we scraped a random subset of 758 accounts that we refer to as the random sample.

A.3 Construction of News Outlets Dataset

To construct the list of news outlets we consider in this study, we begin with a list of 4,412 publishers from Athey et al. [2021] and Watts et al. [2021]. We exclude outlets that publish fewer than 10 articles per month, publishers without a Twitter account, those primarily focused on publishing soft news, and publishers without a slant score in the Robertson et al. [2018] dataset (see below for more detail on this measure). This results in 1,170 hard news publishers on Twitter that are included in the analysis.

For every publisher in this dataset, we assign a slant score capturing how preferred it is by known Democrats or Republicans, as described in Robertson et al. [2018]. Robertson et al. [2018] reports a partisan bias score for 19,022 of the most popular domains. To generate this score, they match voter registration records from around half a million US voters to Twitter accounts, scrape information from over 100 million tweets from the linked accounts over time, identify tweets that contain a URL, and construct a *partisan audience bias score* for each domain, leveraging the sharing propensities of Democrats and Republicans.

Importantly, our slant measure based on Robertson et al. [2018] is normalized between -1 and 1. A completely liberal news outlet has a slant of approximately -1, a moderate news outlet has a slant of approximately 0, and a completely conservative news outlet has a slant of approximately 1.[19]

A.4 Estimates from Fully Saturated Regressions

In the main text, we report the results of a regression that estimates effects separately for the peer information and disclosure conditions following equation (1). The coefficient of interest, β, represents the ITT estimate of receiving peer information (an incentive to disclose the news diet summary) across both types of participants: those in the disclosure (peer information) treatment and those in the disclosure (peer information) control.

Appendix Figure A-13 reports the main estimates of a fully saturated regression in which we simultaneously include participants who received: (1) no peer information and no incentive to disclose the news diet summary, (2) peer information but no incentive to disclose, (3) no peer information but an incentive to disclose, and (4) peer information *and* an incentive to disclose. Panel A reports the results on the probability of updating beliefs about the slant of the peers' news diet. We find that, compared to participants who do not receive peer information or an incentive to disclose (those in group 1), only the groups that receive peer information (groups 2 and 4) differentially update their beliefs. Both of these effects are statistically significant and of a similar magnitude: moving from group 1 to group 2 (4) leads to a 12.74 (12.97) pp increase in the probability of a participant updating their posterior beliefs about their peers. Instead, we find that receiving an incentive to disclose has a quantitatively small and statistically insignificant effect on beliefs across both types

[19]The correlation between this measure and Bakshy et al. [2015] is 0.96. Appendix Figure A-4 presents the distribution of the slant scores of our final dataset of publishers.

of participants: those who do not receive the peer information treatment (group 3 vs. group 1) and those who do (group 4 vs. group 2). Panel B reports the effects on the probability of making changes to the news diet summary (any change). We find that, compared to participants who do not receive peer information or an incentive to disclose (those in group 1), only the groups that receive an incentive to disclose their news diet summary (groups 3 and 4) are more likely to make changes: moving from group 1 to group 3 (4) leads to a 6.22 (4.26) pp increase in the probability of making changes to the news diet. Both of these effects are marginally significant (p-values are 0.086 and 0.124, respectively). We find that the peer condition instead reduces the probability of making changes to the news diet, although these effects are insignificant and smaller in magnitude.

Overall, the interaction term between the two treatment conditions (peer information and disclosure) is quantitatively small and statistically insignificant, which explains why the estimates using this approach are very similar to those in the main text.[20]

A.5 External Validity and Balance Checks

Our sample is more white, more educated, more heavily male, and older than the US adult population (Table 1). This implies that our estimates, while representative of our sample, might not necessarily be representative of the average US person. What do our results imply for the average US adult? While we cannot ultimately rule out the possibility that unobservable traits play an important role, we present two pieces of suggestive evidence in an attempt to approach this question.

First, we estimate an alternative specification of regression equation (1) with sample weights to adjust for these observable differences.[21] Appendix Figures A-14 and A-15 replicate Figures 3 and 9 in the main text but instead using this weighted specification. We observe that, although the magnitude of the treatment effect is smaller in some cases and some of them become noisier,[22] these results closely align with those discussed in the

[20]The magnitude associated with the interaction term is -0.012 (standard error of 0.047) in Panel A and 0.014 (standard error of 0.040) in Panel B.

[21]We calculate the weight of each observation in our sample in two steps. We construct a variable-specific weight so that each observation is weighted according to the share of US adults that have the same trait (e.g. since females are underrepresented in our sample, they receive a higher weight than men). This guarantees that the weighted average in our sample corresponds to an average of US adults. We also rescale each variable so that they sum up to 1. We then create the final weight of each observation as the product of the variables-specific weights, which again is rescaled so that all weights sum up to 1.

[22]The effect of the disclosure treatment on any change is 5.04 pp (standard error of 2.32), which is smaller in magnitude to the respective estimate in the unweighted regression (coefficient of 7.29 pp standard error of 1.58).

main text.

Second, we believe a central trait that explains these demographic differences is that our ads (see Appendix Section A.1) might have attracted politically interested users both because they had political content (a US map about partisan dominance by county) and because some of them incentivized users to complete the study in exchange for learning about their political bias (which might be of particular interest to politically engaged users).[23] In our sample, 60% of people claim to follow politics "very closely."[24] In principle, it is not clear how political engagement should affect the results. While politically engaged individuals tend to be more active, and therefore more likely to be exposed to social interactions,[25] these interactions could be marginally more meaningful and informative for those who are not typically engaged. Appendix Figures A-16 and A-17 replicate Figures 3 and 9 for the sample of participants that did not follow politics very closely (this includes 37% of participants that claim to have followed politics "somewhat closely" and 3% that claim to have followed politics "not at all closely"), which are the under-represented group in our sample. We find that the magnitude of the treatment effects is smaller,[26] but still very salient for this subgroup.

We believe our final sample is of interest in itself as it captures the effects for politically interested participants, who likely engage with political news on a regular basis either directly or through their peers. The results below indicate that while the effects are slightly smaller, they are also present when looking directly at under-represented groups in our sample.

We now turn to balance checks. Appendix Table A-4 reports average covariates (from both the baseline survey and scraped Twitter data) for the disclosure and peer information conditions. We find that characteristics across treatment and control groups in both conditions are very similar consistent with a randomized design.

[23]We find, based on the most recent round of the World Values Survey for the US, that the direction of the divergence in each of these demographic characteristics between our sample and the population of US adults always indicates greater political engagement. Male, white, graduate, and above-median-age individuals are, on average, 14, 6, 16, and 23 pp *more* likely to be interested in politics than female, non-white, non-graduate, and below-median-age individuals, respectively.

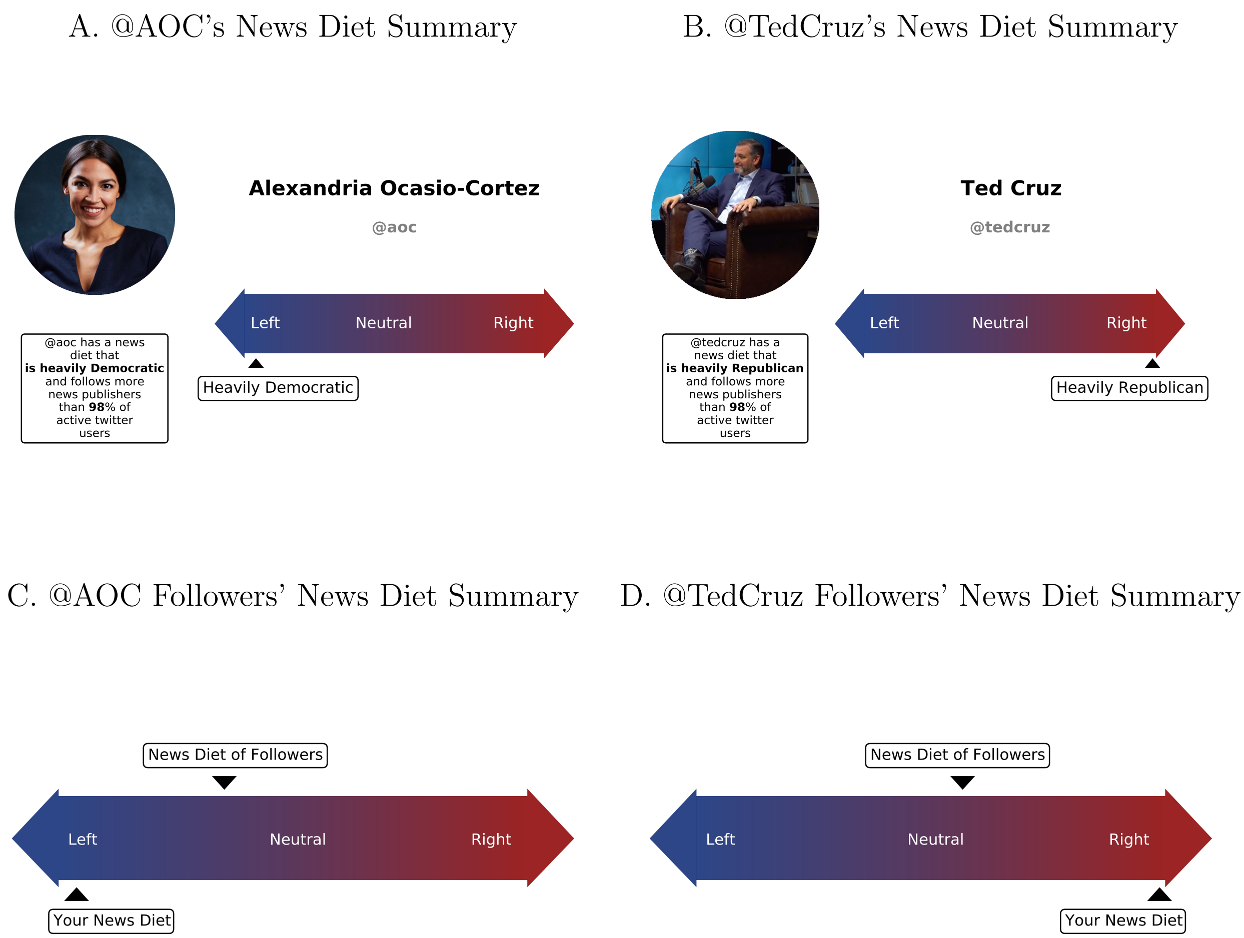
[24]Rather than asking how closely participants follow politics, the World Values Survey asks how interested are individuals in politics. We find that only 20% claim to be very interested in politics. The rest either have no interest at all (14%) or have an intermediate interest (66%).

[25]In our sample, the median participants who follow politics very closely follow 164 more accounts and three more news outlets than the median participants who do not.

[26]The effect of the disclosure treatment on any change is 4.59 pp (standard error of 2.43), which is smaller in magnitude than the respective estimate in the unweighted regression (coefficient of 7.29 pp standard error of 1.58).

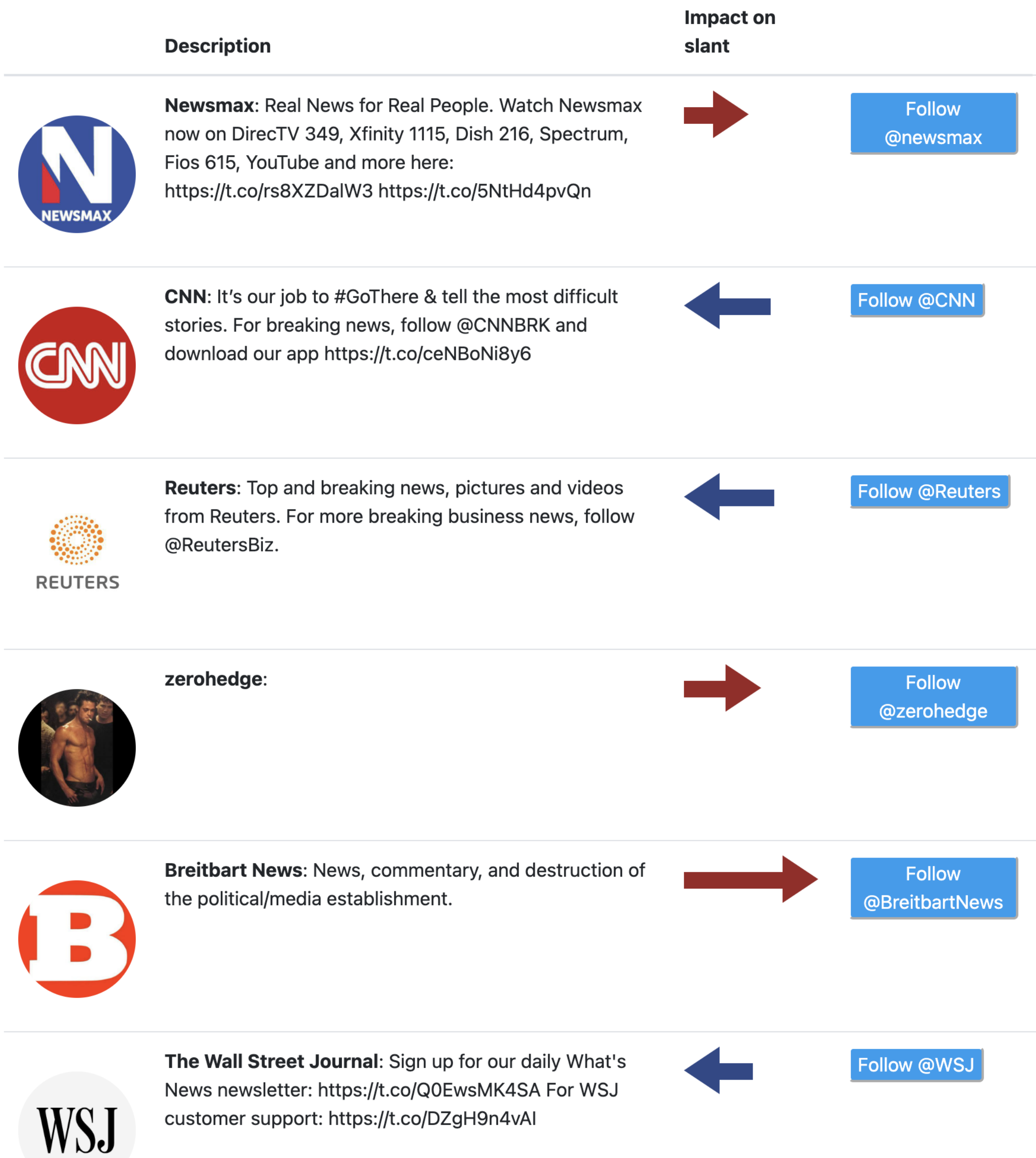
A.6 Additional Figures and Tables

Figure A-1: News Diet Summary Example



Notes: The figure shows the design of the infographics conveying the news diet summary. It includes examples using information of the Twitter accounts of two well-known US politicians. However, study participants received this information about their own Twitter account or that of their followers. All study participants saw an infographic conveying information about their own news diet summary (as in Panels A or B), but only those assigned to the information treatment group viewed additional information about the political ideology of their followers (as in Panels C or D).

Figure A-2: News Outlets Recommendations Example

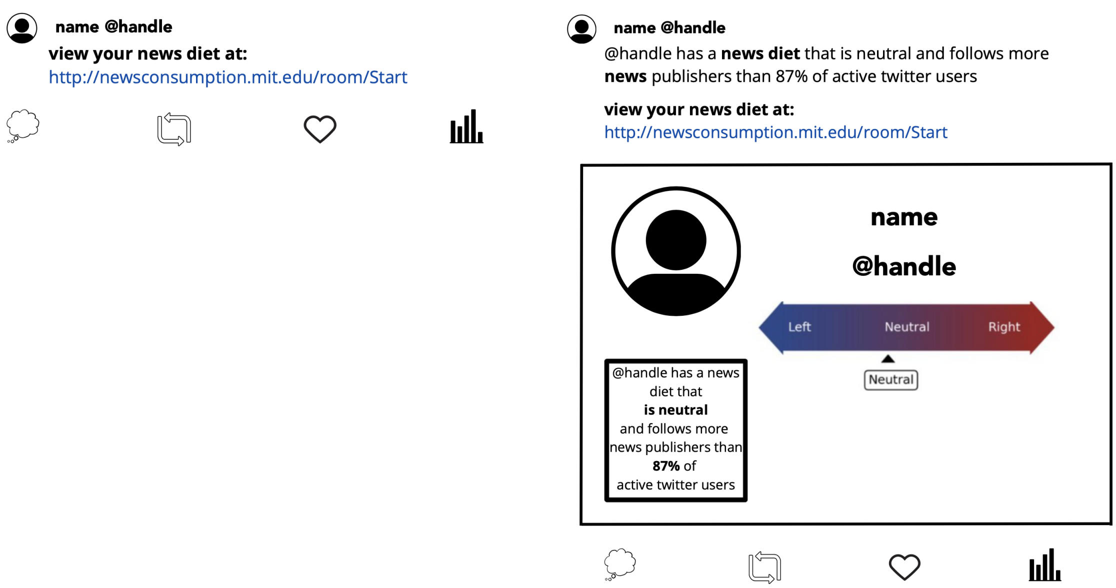


Notes: The figure shows an example of the recommended news outlets that participants received on their recommendation page. Participants visualized six news outlets, three conservative and three liberal. We explained to the participant how following each of these news outlets would affect the slant of their news diet summary.

Figure A-3: Drafted Tweets for Compliance in Disclosure Condition

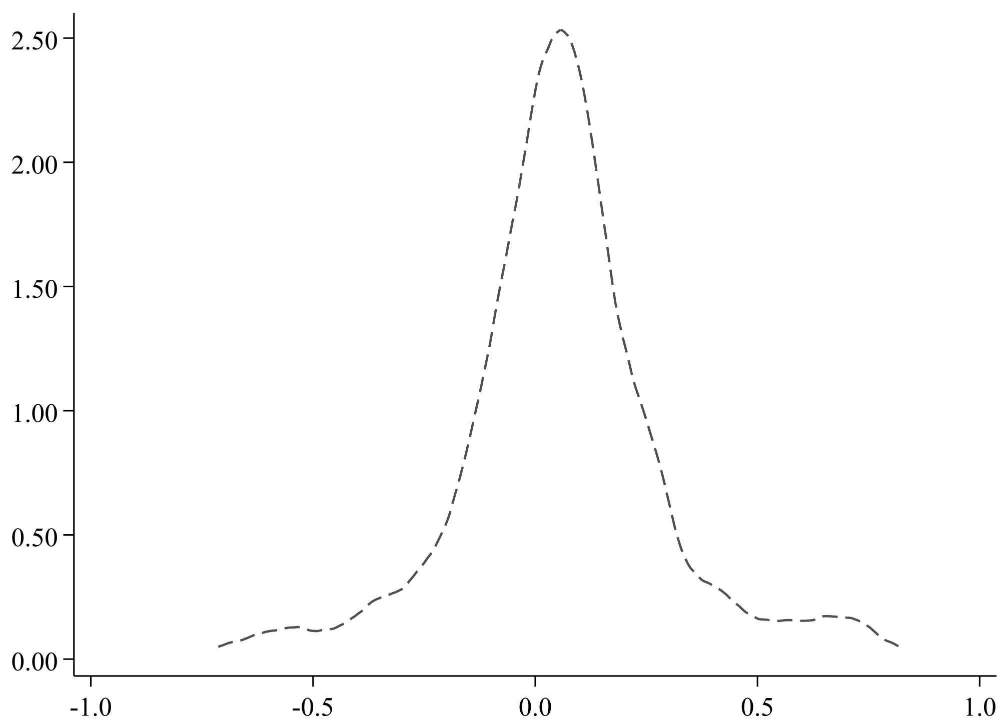
A. Control

B. Treatment



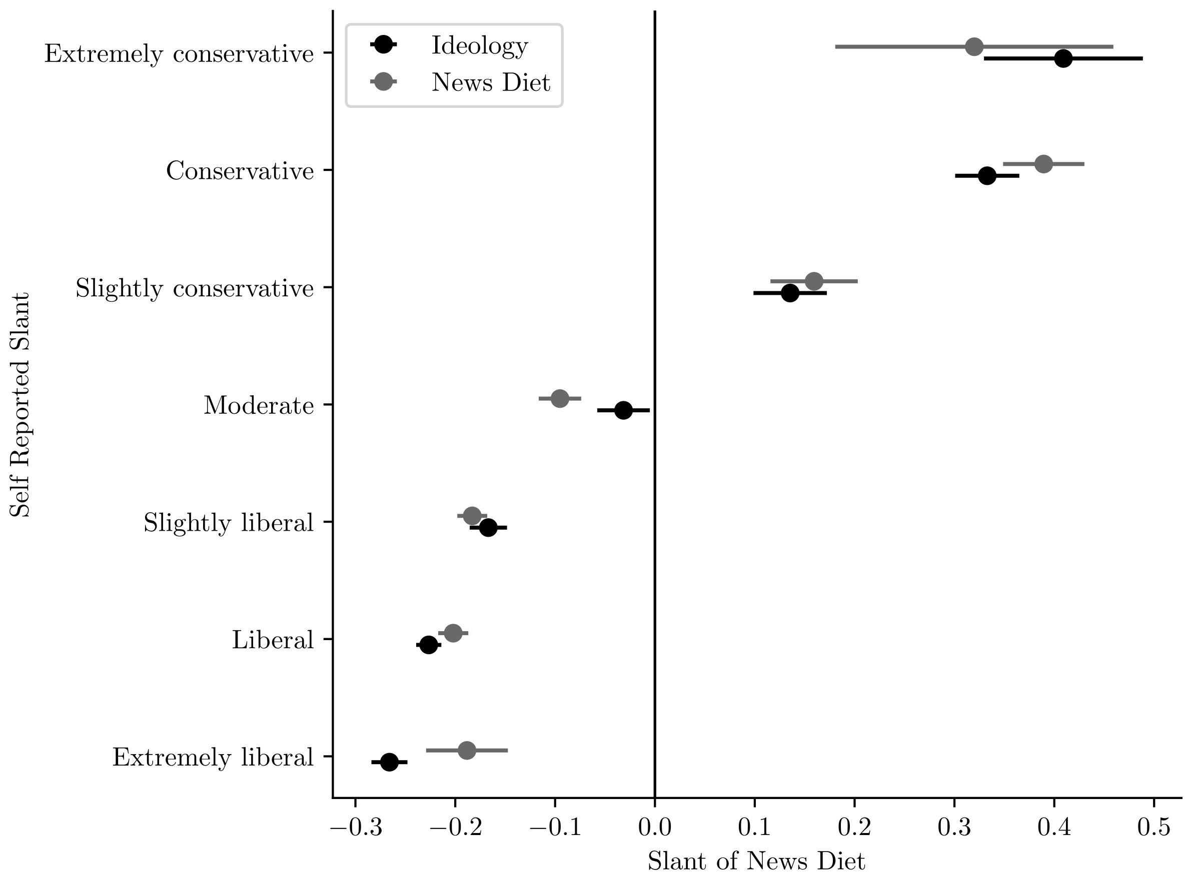
Notes: This figure shows the drafted tweets that participants were incentivized to tweet in the disclosure condition. Participants in the control group were asked to tweet a referral link containing no personal information inviting others to participate in our intervention (left panel). Those in the treatment group were asked to tweet the referral link as well as the infographic containing the news diet summary that has information on both the number of news outlets followed and its slant (right panel).

Figure A-4: Distribution of the Slant of the Twitter News Outlets



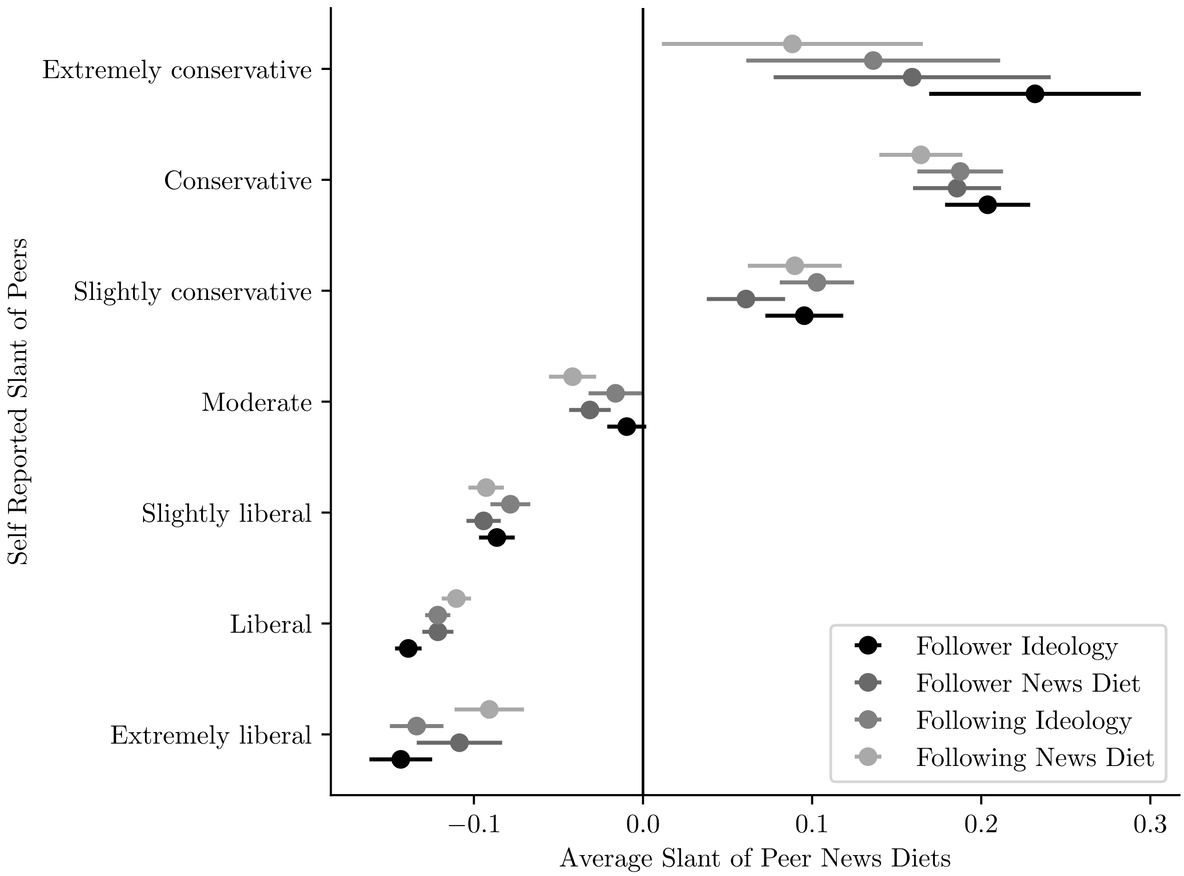
Notes: This figure plots the distribution of the slant for the final dataset of 1,170 news outlets on Twitter.

Figure A-5: Correlation between News Outlets Followed and Self-Reported Data



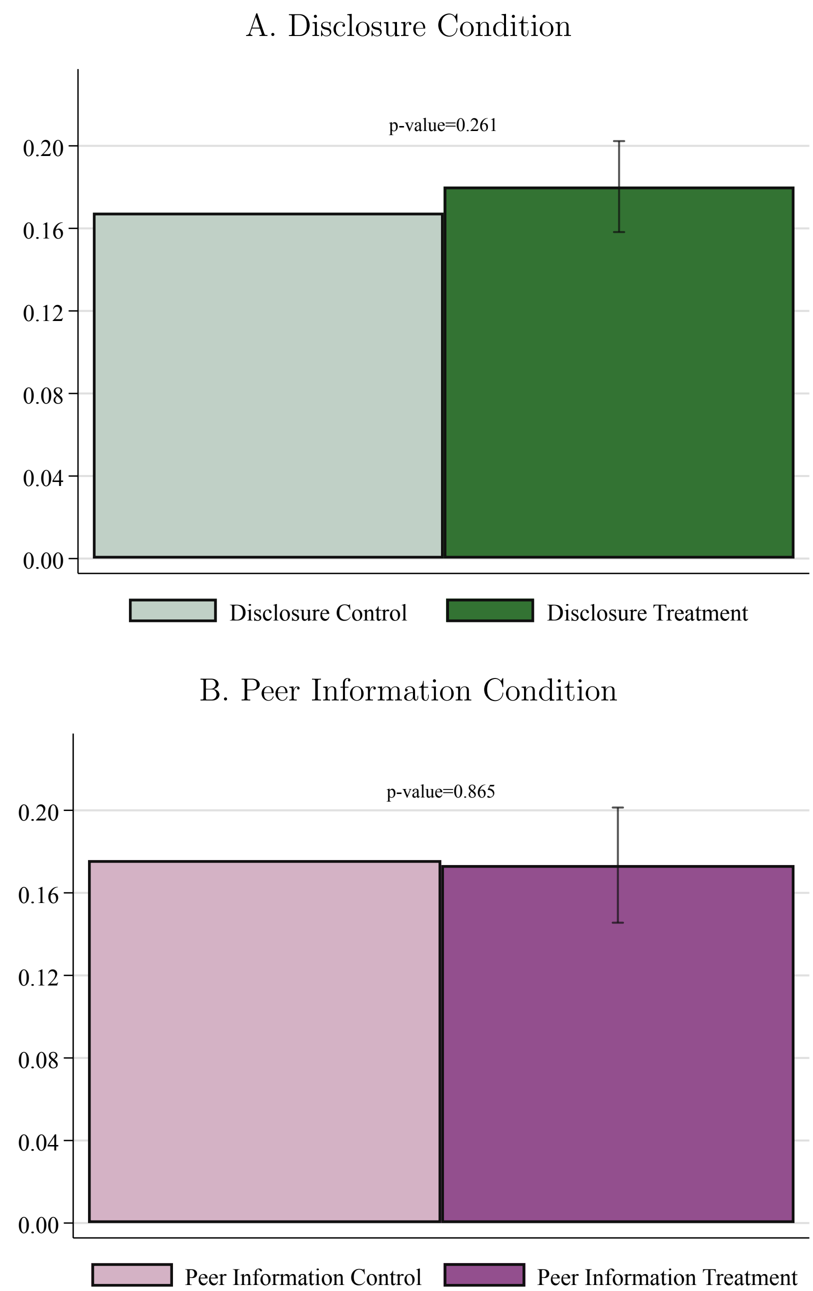
Notes: This figure plots the average slant of users at the start of the experiment by self-reported ideology and self-reported slant of the participant's news diet. The whiskers indicate the 95% confidence intervals of the average slant of users' news diets at the beginning of the experiment.

Figure A-6: Correlation between News Outlets Followed by Peers and Self-Reported Data



Notes: This figure plots the average slant of users' followers by various self-reported measures. Follower (Following) Ideology indicates the self-reported ideology of the individuals who follow the participant (the participant follows) on Twitter. Follower (Following) News Diet indicates the self-reported average slant of publishers followed by individuals who follow the participant (the participant follows) on Twitter. The whiskers indicate the 95% confidence intervals of the average slant of users' news diets at the beginning of the experiment.

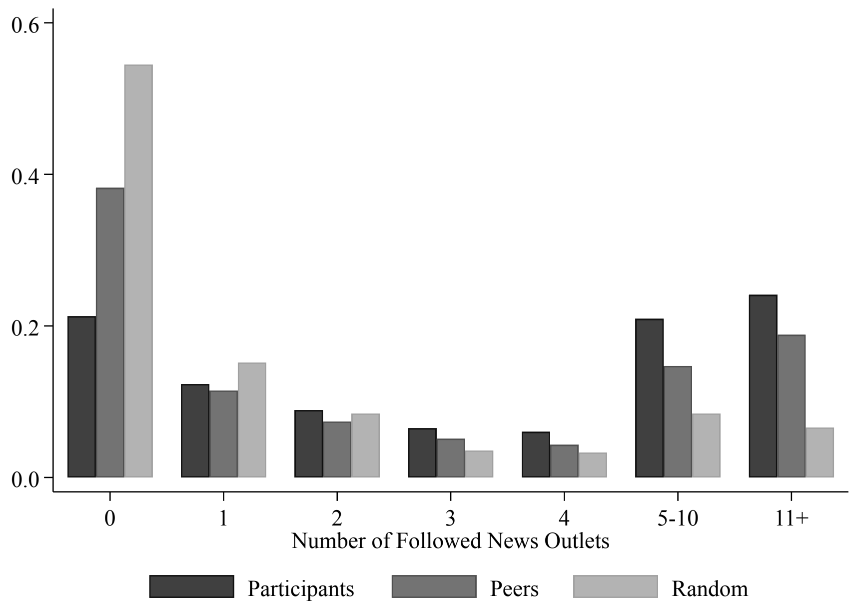
Figure A-7: Differential Attrition Rate for Disclosure and Peer Information Conditions



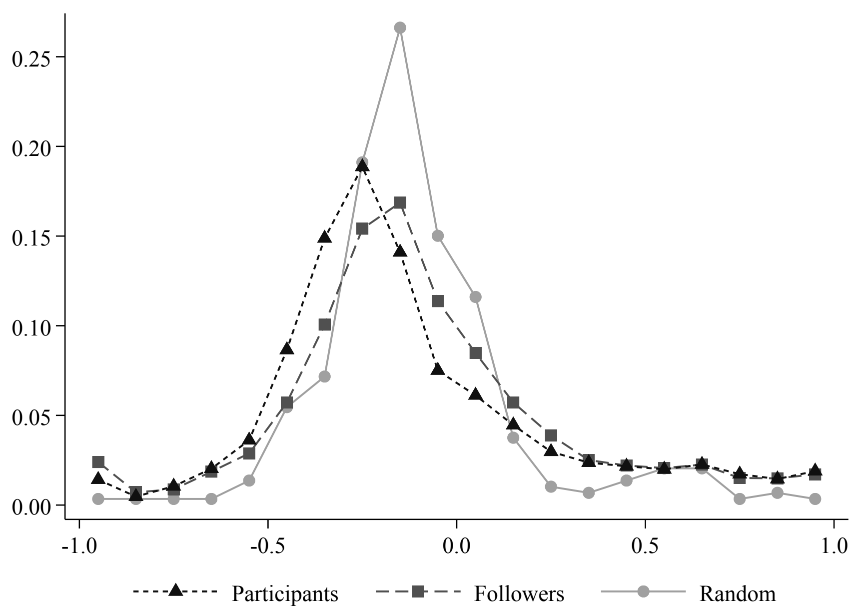
Notes: This figure reports the fraction of individuals that left the experiment before completion across participants assigned to the treatment and control groups in the disclosure condition (Panel A) and the treatment and control groups in the peer information condition (Panel B). The difference between the treatment and control groups in the disclosure condition is that only the former is asked to reveal information about which news outlets they follow to their peers, which is done after they have the chance to modify these outlets (see more details in Section 2.1). The difference between the treatment and control groups in the peer information condition is that only the former receives feedback on the slant of a random sample of their peers/followers (more details are available in Section 2.1). The whiskers indicate the 95% confidence intervals of the difference between the effect in the treatment and control groups following equation (1). We also report the p-value of the null hypothesis of no difference between the treatment and control groups.

Figure A-8: News Outlets Followed

A. Total Number of News Outlets Followed

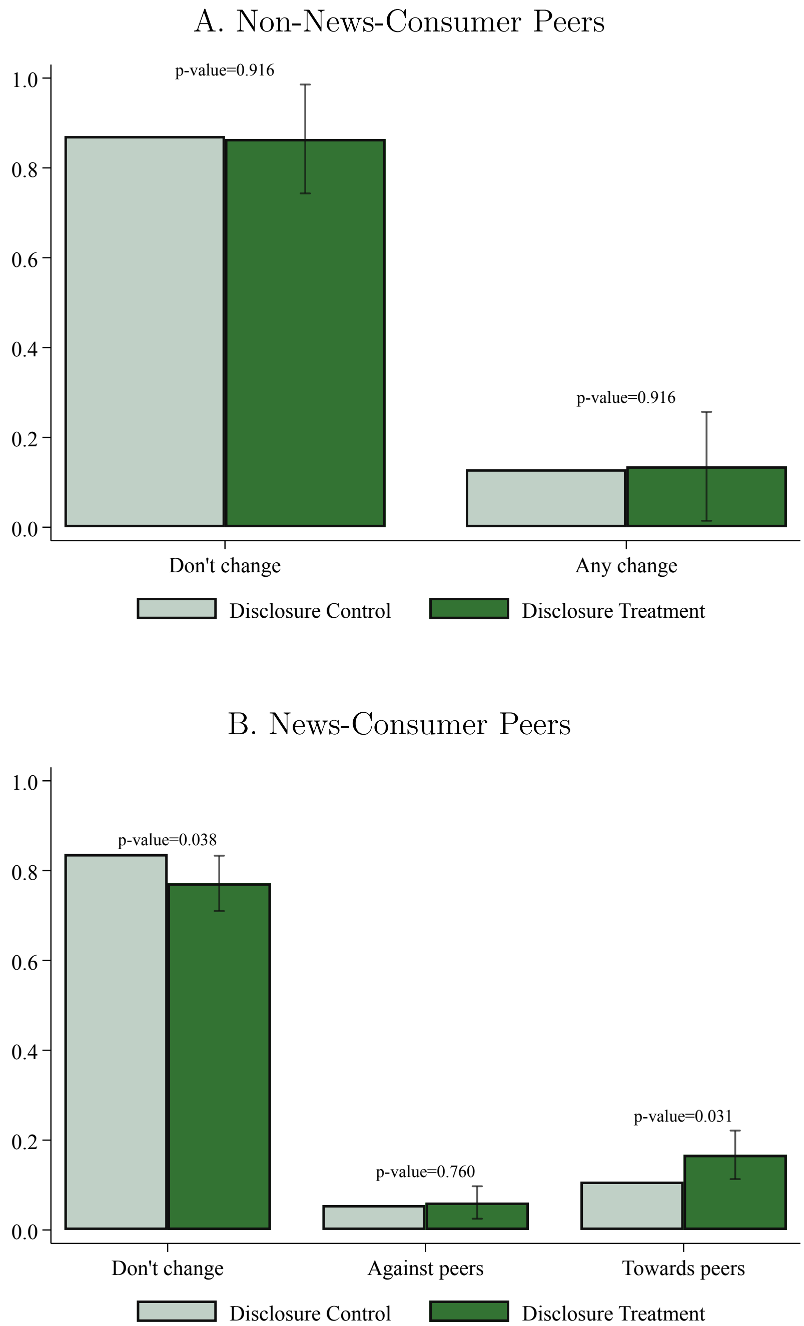


B. Slant of News Outlets Followed Conditional on Following Outlets



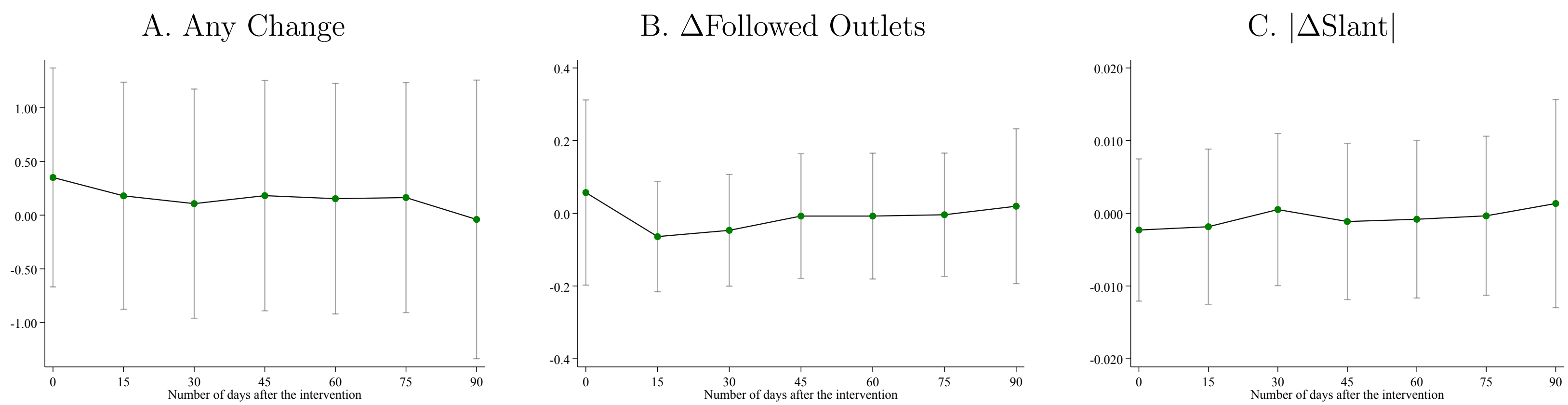
Notes: This figure presents the total number of news outlets followed (Panel A) and the average slant of these outlets (conditional on following at least one outlet) for three samples: the participants, their peers/followers, and a random sample of active Twitter users (see Appendix Section A.2).

Figure A-9: Disclosure Condition and Movements Toward the Center and Peers – Non-News Consumers



Notes: This figure reports the treatment effects of the disclosure condition on whether non-news consumers (participants who do not follow any news outlet at baseline, 20% of our sample) adjust their slant. Panel A reports the results for the group of participants with non-news-consumer peers, while Panel B displays the results for participants with news-consumer peers. The difference between the treatment and control groups in the disclosure condition is that only the former is asked to reveal information about which news outlets they follow to their peers, which is done after they have the chance to modify these outlets (see more details in Section 2.1). The whiskers indicate the 95% confidence intervals of the difference between the effect in the treatment and control groups following equation (1). We also report the p-value of the null hypothesis of no difference between the treatment and control groups.

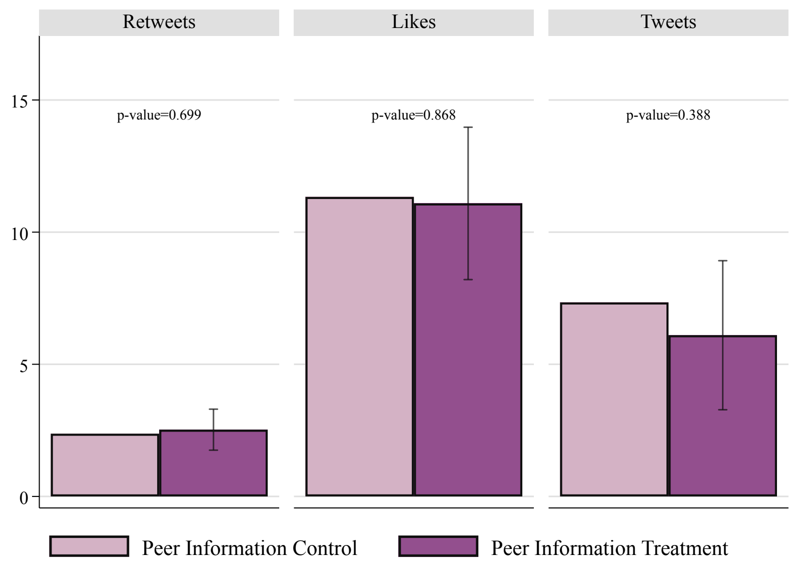
Figure A-10: Peer Information Condition and Treatment Effects Over Time



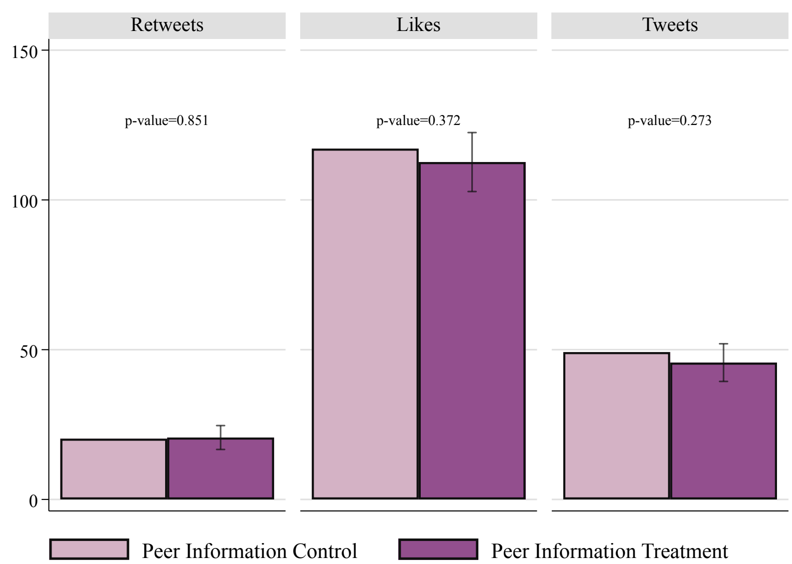
Notes: This figure reports ITT estimates (comparing the treatment and control groups in the peer treatment) on: an indicator variable if the participant makes any change to the news outlets they follow (left), the change in the number of news outlets followed (center), and the change in the absolute slant of the news outlets followed (right). In all cases, we compare these outcomes at different points in time in the post-intervention period (see x-axis) vs. baseline. The difference between the treatment and control groups in the disclosure condition is that only the former is asked to reveal information about which news outlets they follow to their peers, which is done after they have the chance to modify these outlets (see more details in Section 2.1). The whiskers indicate the 95% confidence intervals of the difference between the effect in the treatment and control groups following equation (1).

Figure A-11: Peer Information Condition and Engagement

A. News Engagement



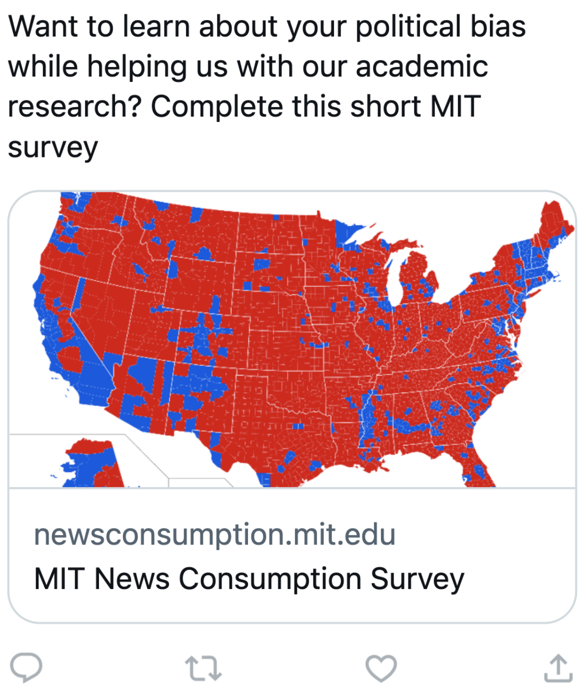
B. Non-News Engagement



Notes: This figure displays the mean of the number of retweets, likes, and tweets associated with news outlets (Panel A) and non-news outlets (Panel B) across participants assigned to the treatment and control groups in the peer condition. Retweets, likes, and tweets are defined as associated with news outlets if the tweet mentions a news outlet or responds to a tweet created by a news outlet. The difference between the treatment and control groups in the peer information condition is that only the former receives feedback on the slant of a random sample of their peers/followers (more details are available in Section 2.1). The whiskers indicate the 95% confidence intervals of the difference between the effect in the treatment and control groups following equation (1). We also report the p-value of the null hypothesis of no difference between the treatment and control groups.

Figure A-12: Recruitment Ads

A. Question/No monetary incentive



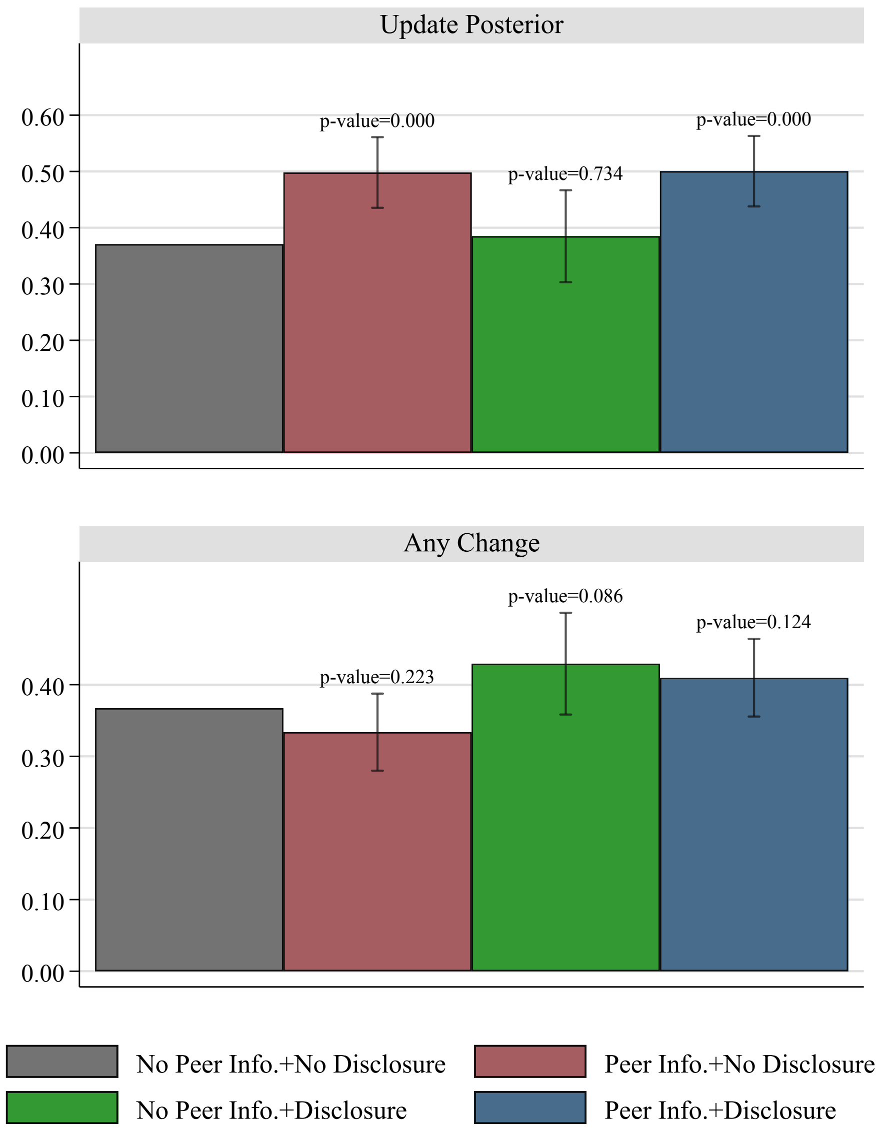
B. Question/Monetary incentive

C. Statement/No monetary incentive

D. Statement/Monetary incentive

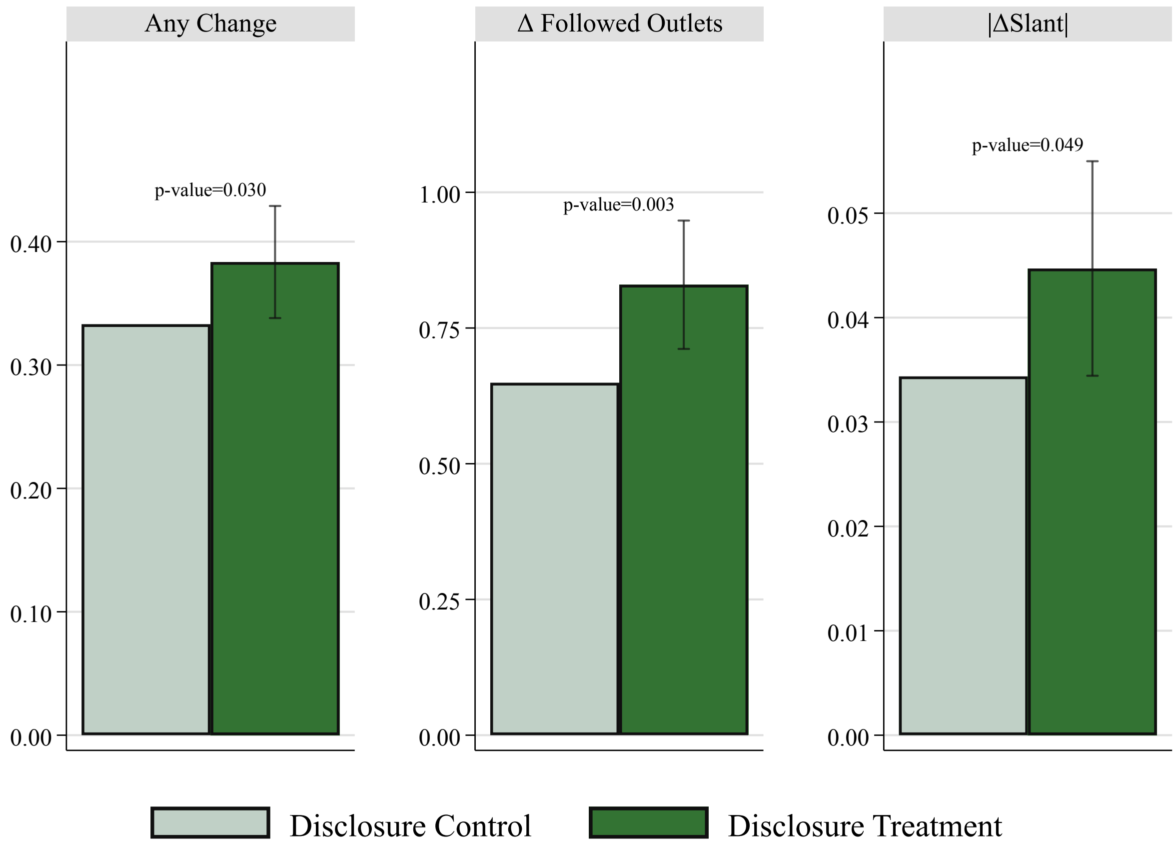
Notes: The figure shows the ads used for recruitment. The maps indicate whether the Republican Party (red) or Democratic Party (blue) obtained a majority of votes in the 2020 presidential election in every US county, and a text that varies either whether a monetary or non-monetary incentive (participants learning about their own political bias) is offered.

Figure A-13: Disclosure and Peer Information Conditions and its Effect on Belief Adjustment and News Outlets Followed



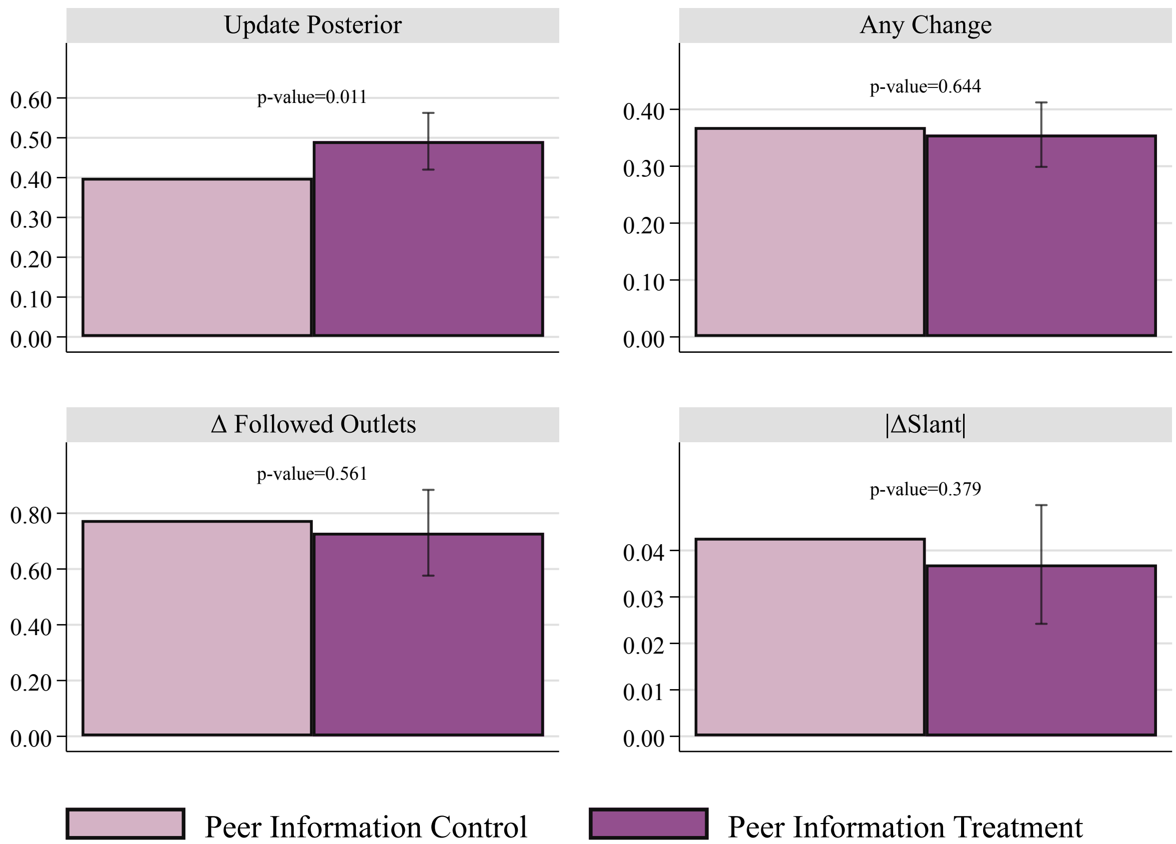
Notes: This figure reports the mean of the following two outcomes across participants assigned to any factorial combination between the treatment and control groups in the peer information and disclosure conditions: an indicator variable if participants update their beliefs regarding the slant of the news consumed by their peers (top), and an indicator variable if a participant makes any change to the news outlets followed (bottom). In all cases, we compare these outcomes in the post-intervention period vs. baseline. The difference between the treatment and control groups in the disclosure condition is that only the former is asked to reveal information about which news outlets they follow to their peers, which is done after they have the chance to modify these outlets (see more details in Section 2.1). The difference between the treatment and control groups in the peer information condition is that only the former receives feedback on the slant of a random sample of their peers/followers (more details are available in Section 2.1). The whiskers indicate the 95% confidence intervals of the difference between the effect in the treatment and control groups following equation (1). We also report the p-value of the null hypothesis of no difference between the treatment and control groups.

Figure A-14: Disclosure Condition and Changes to News Outlets Followed – Weighting on Observable Characteristics



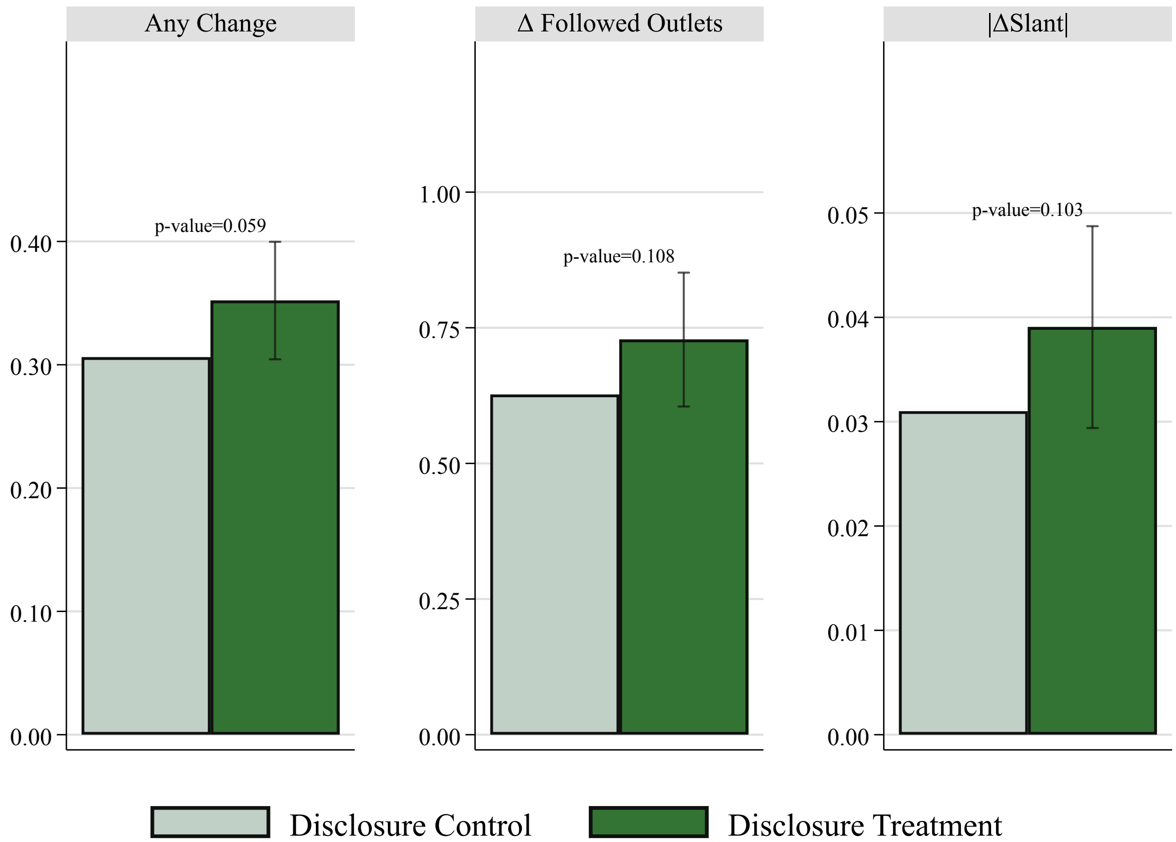
Notes: This figure presents the same results as in Figure 3 but weighting the observations to match the observable characteristics in the US adult population in terms of gender, education, ethnicity, and age. It reports the mean of the following three outcomes across participants assigned to the treatment and control groups in the disclosure condition: an indicator variable if the participant makes any change to the news outlets followed (left), the change in the number of news outlets followed (center), and the change in the absolute slant of the news outlets followed (right). In all cases, we compare this outcome in the post-intervention period vs. baseline. The difference between the treatment and control groups in the disclosure condition is that only the former is asked to reveal information about which news outlets they follow to their peers, which is done after they have the chance to modify these outlets (see more details in Section 2.1). The whiskers indicate the 95% confidence intervals of the difference between the effect in the treatment and control groups following equation (1). We also report the p-value of the null hypothesis of no difference between the treatment and control groups.

Figure A-15: Peer Information Condition, Belief Updating, and News Outlets Followed – Weighting on Observable Characteristics



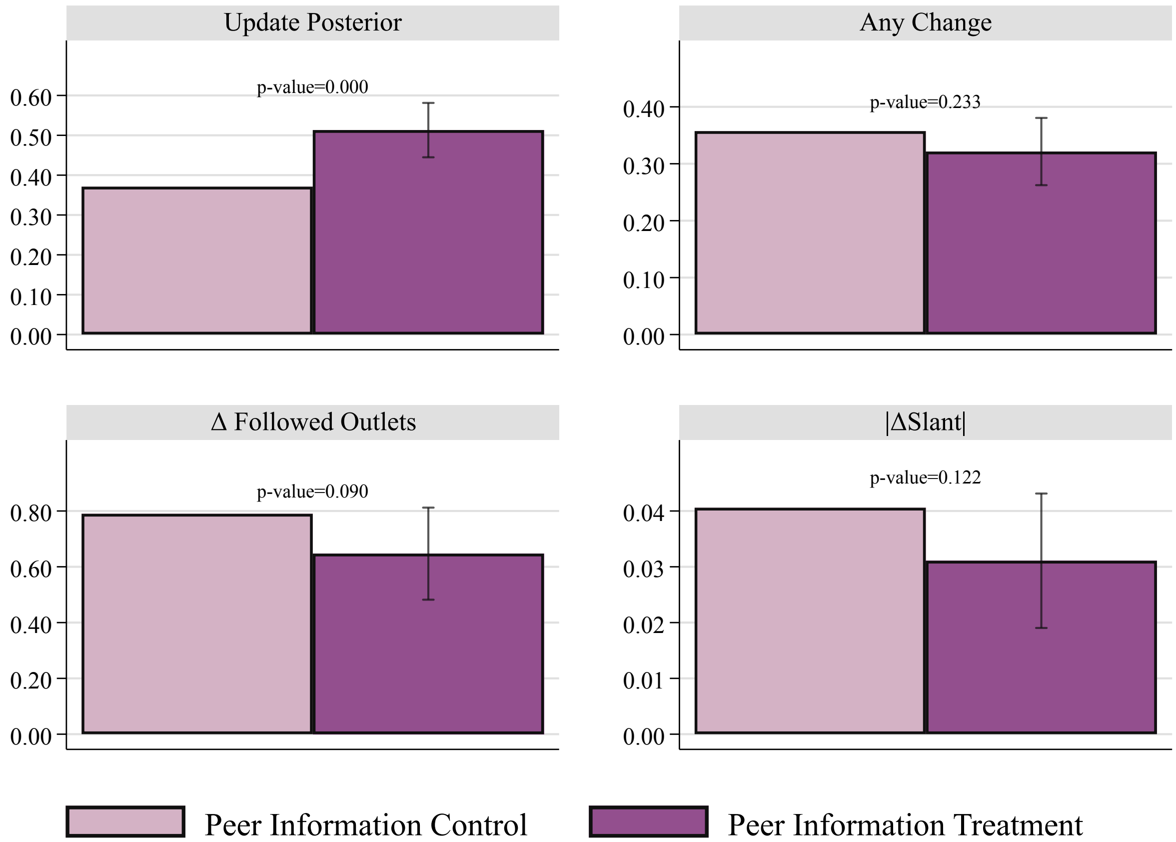
Notes: This figure presents the same results as in Figure 9 but weighting the observations to match the observable characteristics in the US adult population in terms of gender, education, ethnicity, and age. It reports the mean of the following four outcomes across participants assigned to the treatment and control groups in the peer condition: an indicator variable if participants update their beliefs about the slant of the news consumed by their peers (upper left), an indicator variable if a participant makes any change to the news outlets followed (upper right), the change in the number of news outlets followed (bottom left), and the change in the absolute slant of the news outlets followed (bottom right). In all cases, we compare these outcomes in the post-intervention period vs. baseline. The difference between the treatment and control groups in the peer information condition is that only the former receives feedback on the slant of a random sample of their peers/followers (more details are available in Section 2.1). The whiskers indicate the 95% confidence intervals of the difference between the effect in the treatment and control groups following equation (1). We also report the p-value of the null hypothesis of no difference between the treatment and control groups.

Figure A-16: Disclosure Condition and Changes to News Outlets Followed – Participants That Do Not Follow Politics Very Closely



Notes: This figure displays the same results as in Figure 3 for the sample of participants that do not follow politics very closely. It presents the mean of the following three outcomes across participants assigned to the treatment and control groups in the disclosure condition: an indicator variable if the participant makes any change to the news outlets followed (left), the change in the number of news outlets followed (center), and the change in the absolute slant of the news outlets followed (right). In all cases, we compare this outcome in the post-intervention period vs. baseline. The difference between the treatment and control groups in the disclosure condition is that only the former is asked to reveal information about which news outlets they follow to their peers, which is done after they have the chance to modify these outlets (see more details in Section 2.1). The whiskers indicate the 95% confidence intervals of the difference between the effect in the treatment and control groups following equation (1). We also report the p-value of the null hypothesis of no difference between the treatment and control groups.

Figure A-17: Peer Condition, Belief Updating, and News Outlets Followed – Participants That Do Not Follow Politics Very Closely



Notes: This figure presents the same results as in Figure 9 for the sample of participants that do not follow politics very closely. It reports the mean of the following four outcomes across participants assigned to the treatment and control groups in the peer condition: an indicator variable if participants update their beliefs about the slant of the news consumed by their peers (upper left), an indicator variable if a participant makes any change to the news outlets followed (upper right), the change in the number of news outlets followed (bottom left), and the change in the absolute slant of the news outlets followed (bottom right). In all cases, we compare these outcomes in the post-intervention period vs. baseline. The difference between the treatment and control groups in the peer information condition is that only the former receives feedback on the slant of a random sample of their peers/followers (more details are available in Section 2.1). The whiskers indicate the 95% confidence intervals of the difference between the effect in the treatment and control groups following equation (1). We also report the p-value of the null hypothesis of no difference between the treatment and control groups.

Table A-1: Descriptive Statistics about Twitter Engagement

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Sample					Pew Survey	
	Median	Percentile 25	Percentile 75	Median Bottom 10% Tweeters	Median Top 10% Tweeters	Median Bottom 10% Tweeters	Median Top 10% Tweeters
Tweets	24	12	187	6	338	2	138
Followers	148	35	604	12	467	19	387
Followings	500	193	1,107	146	764	74	456

Notes: This table reports summary statistics of the distribution of the number of tweets, followers, and followings in our main sample (Columns 1–5) and a sample of a nationally representative panel of US adults with an active Twitter account (Columns 6–7). For details of this panel, see `https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/`.

Table A-2: Peer Information Condition and Belief Adjustment on Followers and Subsequent Changes to News Choices – Variation in Accuracy of Feedback

	(1)	(2)	(3)	(4)
	C. Normalized Belief Adjustment Followers			
	Belief Change		Sign (Belief Change)	
Peer Treatment	0.088	0.088	0.072	0.073
	(0.018)	(0.017)	(0.018)	(0.018)
Peer Treatment × Distance Feedback vs Truth		0.139		0.100
		(0.043)		(0.041)
	D. Normalized News Choice Adjustment			
	Belief Change		Sign (Belief Change)	
Peer Treatment	-0.008	-0.009	-0.020	-0.021
	(0.019)	(0.019)	(0.017)	(0.017)
Peer Treatment × Distance Feedback vs Truth		0.029		0.044
		(0.026)		(0.036)
Observations	3,757	3,757	3,757	3,757

Notes: This table reports ITT estimates (comparing the treatment and control groups in the peer condition) on normalized belief adjustment about the *followers* (Panel A) and normalized adjustments in the slant of news outlets (Panel B) following equation (1). Belief or slant adjustments are defined as the difference between the post-intervention and the baseline period. Unlike in Table 2, where we normalize belief adjustment by looking at the difference between the feedback and prior beliefs (see equation 3), here we normalize belief adjustments by comparing the difference between the feedback and the truth. For each set of outcomes, we report both the intensive (how much participants move in the direction of the feedback, Columns 1 and 2) and extensive margin (whether participants move in the direction of the feedback, Columns 3 and 4). The even columns present heterogeneous treatment effects by the absolute value of the difference between the prior and the feedback. The outcomes and the distance between the prior and the feedback are both standardized. The difference between the treatment and control groups in the peer information condition is that only the former receives feedback on the slant of a random sample of their peers/followers (more details are available in Section 2.1). Standard errors reported in parentheses are robust against heteroscedasticity.

Table A-3: Peer Information Condition and Belief Adjustment on Followings and Subsequent Changes to News Choices – Exogenous Variation on the Accuracy of the Feedback

	(1)	(2)	(3)	(4)
A. Peer Treatment Group Sample, Beliefs	Belief Adjustment – Followings			
	Belief Change		Sign (Belief Change)	
Feedback > Truth	0.128	0.129	0.123	0.124
	(0.018)	(0.018)	(0.019)	(0.019)
(Feedback > Truth) × Network Connectedness		0.059		0.070
		(0.024)		(0.025)
B. Peer Treatment Group Sample, News Choices	Normalized Adjustment in Slant of Outlets			
	Slant Change		Sign (Slant Change)	
Feedback > Truth	-0.003	-0.002	-0.005	-0.004
	(0.019)	(0.019)	(0.019)	(0.019)
(Feedback > Truth) × Network Connectedness		0.009		0.012
		(0.028)		(0.027)
Observations	2,953	2,953	2,953	2,953
C. Peer Control Group Sample, Beliefs	Belief Adjustment – Followings			
	Belief Change		Sign (Belief Change)	
Feedback > Truth	0.038	0.040	0.057	0.059
	(0.037)	(0.037)	(0.039)	(0.039)
(Feedback > Truth) × Network Connectedness		-0.062		-0.058
		(0.056)		(0.055)
D. Peer Control Group Sample, News Choices	Normalized Adjustment in Slant of Outlets			
	Slant Change		Sign (Slant Change)	
Feedback > Truth	0.008	0.010	0.035	0.035
	(0.037)	(0.037)	(0.037)	(0.038)
(Feedback > Truth) × Network Connectedness		-0.052		0.005
		(0.045)		(0.049)
Observations	722	722	722	722

Notes: This table reports estimates from equation (2) on beliefs about followings (Panels A and C) and slant adjustments (Panels B and D), defined as the difference between the post-intervention and the baseline period. Panels A and B (C and D) report the results for the sample of participants assigned to the treatment (control) group in the peer condition. *Feedback > Truth* is an indicator variable equal to one if the slant of news outlets followed by peers is greater (to the right) when constructed based on a random sample of five peers (the feedback) vs. a larger sample of peers (the truth, see Section 2.4). For each set of outcomes, we report both the intensive (how much participants move to the right, Columns 1 and 2) and extensive margin (whether participants move to the right, Columns 3 and 4). The even columns present heterogeneous treatment effects by the absolute value of the difference between the feedback and the truth. The outcomes and the distance between the feedback and the truth are both standardized. The difference between the treatment and control groups in the peer information condition is that only the former receives feedback on the slant of a random sample of their peers/followers (more details are available in Section 2.1). Standard errors reported in parentheses are robust against heteroscedasticity.

Table A-4: Balance across Disclosure and Peer Information Conditions

	(1)	(2)	(3)	(4)	(5)	(6)
	Disclosure Condition			Peer Information Condition		
	Control	Treatment	p-value	Control	Treatment	p-value
A. Baseline Survey						
Male	0.695	0.669	0.088	0.679	0.683	0.835
Age	50.639	50.362	0.544	49.825	50.666	0.143
White	0.901	0.911	0.299	0.880	0.913	0.012
Graduate Degree	0.495	0.485	0.538	0.484	0.491	0.741
Ideology	3.054	2.989	0.251	2.985	3.031	0.520
Followers Ideology	3.074	3.036	0.433	3.081	3.049	0.610
Followings Ideology	3.206	3.156	0.294	3.155	3.187	0.586
News Diet	3.287	3.231	0.205	3.239	3.265	0.655
Followers News Diet	3.294	3.285	0.836	3.250	3.299	0.383
Followings News Diet	3.229	3.201	0.548	3.201	3.219	0.768
B. Twitter Data						
Slant	-0.089	-0.095	0.630	-0.094	-0.092	0.890
Peer Slant	-0.047	-0.054	0.246	-0.051	-0.050	0.978
Number of Followers	1,230.010	1,044.122	0.476	1,096.042	1,148.120	0.867
Number of Followings	947.509	1,050.458	0.041	1,043.603	987.567	0.418
Number of Publishers	7.517	8.357	0.095	7.692	7.991	0.568

Notes: This table reports average covariates by treatment status. Panels A and B report covariates based on the baseline survey and Twitter data, respectively. The difference between the treatment and control groups in the disclosure condition is that only the former is asked to reveal information about which news outlets they follow to their peers, which is done after they have the chance to modify these outlets (see more details in Section 2.1). The difference between the treatment and control groups in the peer information condition is that only the former receives feedback on the slant of a random sample of their peers/followers (more details are available in Section 2.1). We also report the p-value of the null hypothesis of no difference between the treatment and control groups.