ISSN 2279-9362

CCA

Sharing News Left and Right: The Effects of Policies Targeting Misinformation on Social Media

Daniel Ershov
Juan S. Morales

No. 651
May 2021

Carlo Alberto Notebooks

www.carloalberto.org/research/working-papers

Sharing News Left and Right: The Effects of Policies Targeting Misinformation on Social Media*

Daniel Ershov† Juan S. Morales‡

April 28, 2021

Abstract

We study Facebook's and Twitter's policy interventions which aimed to reduce the spread of misinformation during the 2020 US election. Facebook changed its news feed algorithm to reduce the visibility of content, while Twitter changed its user interface, nudging users to be thoughtful about sharing content. Using data on tweets and Facebook posts published by news media outlets, we show both policies significantly reduced news sharing, but the reductions varied heterogeneously by outlets' factualness and political slant. On Facebook, content sharing fell relatively more for low-factualness outlets. On Twitter, content sharing fell relatively more for left-wing and high-factualness outlets.

JEL Codes: D72, L82, L86, O33

Keywords: social media; news sharing; media slant; fake news; misinformation

*We thank Charles Angelucci, Julia Cagé, Andrew Guess, Krzysztof Krakowski, Ignacio Monzón and Andrey Simonov for comments and suggestions. Edoardo Bella provided excellent research assistance. Daniel Ershov would like to acknowledge support received from ANR under grant ANR-17-EUR-0010 (Investissements d'Avenir program). All errors are our own.

†Toulouse School of Economics. Address: Université Toulouse 1 Capitole, 1, Esplanade de l'Université, 31080 Toulouse, France. E-mail: daniel.ershov@tse-fr.eu.

‡Collegio Carlo Alberto and ESOMAS Department, University of Turin. Address: Piazza Arbarello 8, Torino, TO, 10122, Italy. E-mail: juan.morales@carloalberto.org.

1 Introduction

Online social media platforms such as Facebook, Twitter and YouTube serve as information gatekeepers and news aggregators. Social media is also associated with increased political polarization, the emergence of information silos or "echo chambers", and the spread of misinformation and fake news (Allcott and Gentzkow, 2017; Zhuravskaya, Petrova and Enikolopov, 2020; Levy, 2021). These issues present social media platforms (and potential regulators) with the challenges of moderating online content to limit the harms of misinformation.

In this paper we study two policy interventions, by Facebook and Twitter, which aimed to "limit the spread of misleading information" ahead-of and during the 2020 US election (APnews.com). Both policies introduced friction to the process of online information diffusion, but they had important differences. Twitter's policy intervention, introduced on October 20 2020, was user-centric and could be characterized as being "bottom-up". Twitter modified its user interface (UI) by changing the default functionality of the sharing button (the retweet) to prompt users to add a comment on the content they wanted to share (TechCrunch.com). Facebook's policy intervention, introduced around election day, November 3 2020, was "top-down." Facebook altered its algorithms to "down-rank... some posts in users' Facebook and Instagram feeds" thereby reducing the visibility (and sharing) of political news (Engadget.com).

We use 284,000 tweets and 197,000 Facebook posts published by 137 popular US media outlets around the time of the US 2020 election to study these policies. We first use an interrupted-time-series empirical strategy to estimate short-run changes in content sharing, as measured by retweets/shares, in the 30-days after the policy changes, relative to the 30-days before. We then use short-run difference-in-differences exercises to analyze whether the impacts of the policies were heterogeneous depending on the factualness and political slant of the media outlets.

We show that both policy interventions significantly reduced news sharing. Specifically, content sharing for our sample of media outlets fell by 18 percent on Twitter and by 13 percent on Facebook. Our estimates for Twitter are in line with the company's own announcement of their policy's impact, published in December 2020 (Twitter.com).[1] To the best of our knowledge, Facebook has not provided similar official estimates on the effects of their policy.

Despite comparable average reductions in content sharing, we document signifi-

[1] Evidence of such UI frictions successfully reducing sharing is also found in Henry ⓡ, Zhuravskaya ⓡ and Guriev (2020).

cant differences in how these varied by outlets' characteristics. Facebook's intervention reduced the sharing of content published by outlets with low or mixed factual reporting by 19 percent, significantly more than the sharing of content published by high-factualness outlets, which decreased by 8 percent. The policy also reduced the sharing of content produced by ideologically extreme outlets, but sharing of content produced by politically centrist outlets did not statistically change. Twitter's intervention produced sharply contrasting outcomes. After the UI change, retweeting of content published by high-factualness outlets decreased significantly more relative to sharing for mixed/low-factualness outlets. Twitter's policy also produced uneven outcomes by outlets' political slant. Sharing of content produced by left-wing outlets decreased by 23 percent, while the decrease for right-wing outlets was of only 10 percent. These results suggest that Twitter's policy failed to substantially reduce the spread of misinformation relative to more accurate news, and failed to reduce the spread of politically polarizing content relative to more centrist content, while having significant effects on content sharing across the platform. This interpretation is consistent with Twitter reversing its policy in December 2020.[2]

Outlets' political slant and factualness are significantly correlated. On average, right-wing outlets are less factual in their reporting. We observe this relationship in ratings from independent providers, but similar findings have been documented in previous work (see for instance Faris et al., 2017; Guess, Nyhan and Reifler, 2020). The interaction between political slant and factualness helps explain the differences in effects between Twitter's and Facebook's policies. "Horserace" regressions including both interactions reveal the most important dimension of heterogeneity for each policy: factualness for Facebook, and political slant for Twitter. On Facebook, the heterogeneous change in sharing by outlet political slant becomes statistically insignificant conditional on changes by outlet factualness. In contrast, conditional on outlet political slant, the effect of Twitter's UI change did not vary by outlet factualness. Our findings suggest that conservative Twitter users were less responsive to the platform's user-centric intervention.

The data does not allow us to disentangle the precise mechanisms through which these differences in sharing patterns arose. Nonetheless, previous literature documents that conservatives are more engaged on social media and more likely to share misleading content. Related literature and surveys also document that conservatives are both less likely to trust social media platforms and to believe that the platforms favour the

[2]To the best of our knowledge, there have been no changes to Facebook's policy.

views of liberals over conservatives (PewResearch.org). Right-leaning users may have been more likely to ignore Twitter's prompts than left-leaning users. Facebook's policy relied less on user agency and could better target the sharing of "misinformation" without generating substantially heterogeneous effects along political ideology.

Our findings have two substantial policy implications. First, social media platforms hoping to improve information quality should take into account their interventions' potential heterogeneous effects. Interventions with asymmetric political impact could fail to achieve the intended effects and produce unintended consequences. Second, our results highlight the power that social media platforms can have over the diffusion of information on their platforms. Even Facebook's relatively more successful policy reveals that platforms' subtle algorithmic tweaks can substantially affect the speed and patterns of information transmission. Governments considering regulating content on social media should also be aware of this.

2 Background

Social media platforms primarily feature user generated content and user actions are crucial to the spread of information. On both Twitter and Facebook, users see feeds of posts from other users that they follow. If a given user wants her followers to see a particular post, she retweets/shares it. This interface, combined with the networked structure of the platforms generates information cascades and "viral" content. Previous research shows that misinformation and fake news spread quickly on social media platforms (Vosoughi, Roy and Aral 2018), and that fake news websites rely disproportionately on traffic from social media, with each "share" resulting in up to 20 site visits (Allcott and Gentzkow, 2017). Of particular concern is the spread of misinformation or misleading information about political candidates and election results (Zhuravskaya, Petrova and Enikolopov, 2020).

Social media companies responded to concerns about the spread of misinformation around the 2020 US election with policies directly targeting content sharing.[3] Two weeks ahead of the 2020 US presidential election, Twitter changed the default way in which its sharing button (the *retweet*) worked, hoping to limit the spread of misinformation on its

[3]The platforms had other policies around the 2020 US election that attempted to reduce polarization and misinformation. These policies were common across the two platforms. Both Facebook and Twitter flagged contentious or unverified content as "disputed," including posts by Donald Trump (Reuters.com). Recent research suggests that such policies are effective at reducing misinformation online (Pennycook et al. 2021; Henry ⓡ, Zhuravskaya ⓡ and Guriev 2020). Both Facebook and Twitter also limited political advertising before the election (CNBC.com).

platform. This user interface (UI) change added "friction" to retweets, by popping out the "quote tweet" window instead, encouraging users to add their own comments to the content (Twitter.com, TheVerge.com).

Unlike Twitter, Facebook's response did not primarily target user behaviour but shared content visibility. Facebook reportedly had "break-glass" election measures, which Facebook's Head of Global Affairs described as "effectively throw[ing] a blanket over a lot of content that would freely circulate on our platforms" (USAToday.com). These measures reportedly demoted content on news feeds to limit political news-sharing. After the election, Facebook revealed they resorted to these measures and "down-rank[ed] some posts in users' Facebook and Instagram feeds" (Engadget.com).

3 Data

Our main sample focuses on 137 popular media outlets for whom we collected Twitter and Facebook data, and data on their political slant and factualness. Our measures of political slant and factualness are at the outlet-level. Information about the political slant of media outlets comes from AllSides, a media rating aggregator. Allsides categorizes the political slant/bias of an outlet as one of five ratings: left, lean left, center/mixed, lean right, and right. The ratings are based on blind surveys and editorial content evaluation. Information about factualness comes from Media Bias Fact Check (MBFC), an independent data provider that ranks the factualness of news outlets and adheres to the International Fact-Checking Network's code of principles (Poynter.org). We collected this data from Baly et al. (2018, 2020) who scraped and aggregated MBFC data in 2018 and 2020. MBFC categorizes outlets as high-factualness, mixed-factualness or low-factualness based on editorial fact-checking of outlets' articles. As we have a small number of low-factualness outlets, we group them into a common mixed/low-factualness group in our empirical analyses. Additional details on both sources, including full lists of outlets by political slant and factualness and a discussion of alternative measures are in Appendix A.1.[4]

Tweets were collected using manually coded Twitter handles of outlets through the public Twitter API between November 10, 2020 and January 18, 2021. Facebook posts

[4]In addition to the 137 outlets for whom we collected factualness and political slant data, we collected AllSides political slant and Twitter data for 192 additional outlets, and slant and Facebook data for 153 additional outlets. Since some specifications of our empirical analysis do not require factualness data, we use the larger sample as a robustness check. Estimates are qualitatively and quantitatively similar across the "main" and "extended" samples for these specifications.

were collected using CrowdTangle.com in March 2021.[5] For each tweet/Facebook post, we collected the full text of the post, the number of times it was retweeted/shared, and other measures of user engagement: likes for Twitter; likes, comments and other reactions ("wow," "haha," "sad," "angry" and "care") for Facebook.[6] Our main estimating sample includes all tweets and Facebook posts 30 days before and 30 days after each respective policy intervention and consists of approximately 284,000 tweets and 197,000 Facebook posts. Summary statistics for these samples are in the Appendix.

4 Empirical analysis and results

We begin our study with a graphical analysis of news sharing patterns around the time of the policy changes. Our analysis is at the news item (tweet/post) level. We regress the log of retweets/shares on a set of individual day fixed effects from 30 days before to 30 days after each policy intervention, as well as additional controls and media outlet fixed effects (see below).[7] We show estimates of the time fixed-effects in Figure 1, as well as fitted polynomials before and after the implementation of the policies. The left-hand panel shows results for Twitter and the right-hand panel shows results for Facebook. Conditional on our set of controls, we find no strong time trends in retweeting/sharing before the policies were implemented. However, the outcome variable changes sharply at the time of the policies' implementations, and the changes are persistent.[8]

To assess the magnitude and significance of these changes, and given the lack of persistent pre- or post- policy trends for both Facebook and Twitter, we use an interrupted time-series regression to capture the average effects of the policies on retweeting/sharing. Our baseline specification for item i of outlet o published on day t is:

$$log(shares)_{iot} = \alpha Post_t + X_i'\beta + \delta_o + \epsilon_{iot} \quad (1)$$

[5] CrowdTangle is a public insights tool owned and operated by Facebook (CrowdTangle.com).

[6] Both retweets/shares and other measures of engagement were recorded at the time of collection rather than on the day they were posted. However, the lifespan of posts on Facebook and Twitter is very short. Industry observers estimate an average engagement "half-life" (i.e., length of time it takes content to reach 50 percent of its total lifetime engagement) of 18 minutes on Twitter (moz.com) and 30 minutes on Facebook (epipheo.com).

[7] To account for tweets and Facebook posts that were not shared, and following previous work (Cagé, Hervé and Viaud, 2020; Morales, 2021), we define the outcome variable as $log(shares + 1)$.

[8] Although Facebook's policy coincides with the US election, it is unlikely that the decrease in Facebook sharing was caused by the election itself. Due to a historically large number of mail-in ballots, votes were not counted for weeks amid continuing political controversies. It is reasonable to expect that political news sharing/retweeting should increase after election day. The left panel of Figure 1 shows that this was the case for Twitter, although retweets soon fell back to its post-Twitter policy levels.

$Post_t$ is a dummy equal to 1 after the platform's policy activates (October 20 for Twitter and November 3 for Facebook). δ_o are media-outlet fixed effects that control for outlet-specific unobservables. We include tweet/post-level controls (X_i) to capture item characteristics: the length of the tweet/post, whether it contains a hashtag, an @ mention or a url, and dummies for whether tweets are replies or retweets.[9]

We also control for item engagement: the number of likes on Twitter, and the number of responses (likes, comments and other reactions) on Facebook. Engagement allows us to proxy for many item-level unobservable characteristics such as tweet/post quality and compare the sharing of similar items over time, reducing omitted variable bias and increasing precision. However, both engagement and the number of retweets/shares are equilibrium quantities affected by platform policies. A policy slowing information transmission is likely to also reduce likes and other forms of engagement. In other words, engagement controls could introduce a simultaneity bias. Nonetheless, these controls capture important tweet/post heterogeneity and account for the increased activity on social media due to the election. Since likes and retweets are positively correlated and both should decrease following the policy change, simultaneity would bias α towards zero, meaning that we estimate lower bounds of the true effects of the policies.[10]

Estimates of this regression with outlet clustered standard errors are in Column (1) of Table 1. After the policy changes, we find that content sharing decreased by 17.8 percent on Twitter and by 12.5 percent on Facebook. Both estimated changes are statistically significant at the 99% confidence level.

We next test whether the average effect is heterogeneous along outlets' factualness, a measure of the accuracy of their reporting (see 3). Given that the stated goal of both policies was to reduce the spread of misinformation, an effective policy should differentially reduce the sharing of content from outlets whose reporting is less accurate. We estimate the following regression:

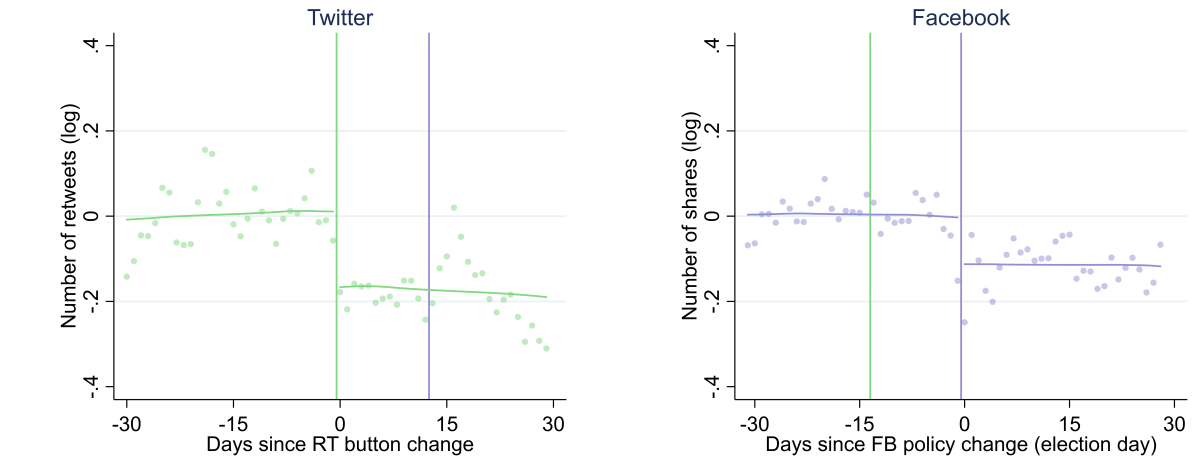
$$log(shares)_{iot} = \alpha_0 Post_t + \alpha_1 Post_t \times \text{High-Factualness}_o + X_i'\beta + \delta_o + \epsilon_{iot} \qquad (2)$$

where $\text{High-Factualness}_o$ is a dummy equal to 1 if outlet o is a high-factualness outlet. α_1 captures the differential effect of the policy for high-factualness outlets relative to

[9]Excluding media outlet retweets from the estimating sample does not change our results.

[10]Omitted variables, in contrast, could bias the coefficients of interest upwards or downwards relative to the true effects. We produce qualitatively similar results without controlling for engagement in the Appendix. The average sharing decrease for Twitter in these regressions is approximately half of the baseline estimate and of Twitter's own estimate of the policy impact (Twitter.com), suggesting the importance of controlling for tweet/post heterogeneity.

Figure 1: Social media platforms' policy changes and news sharing



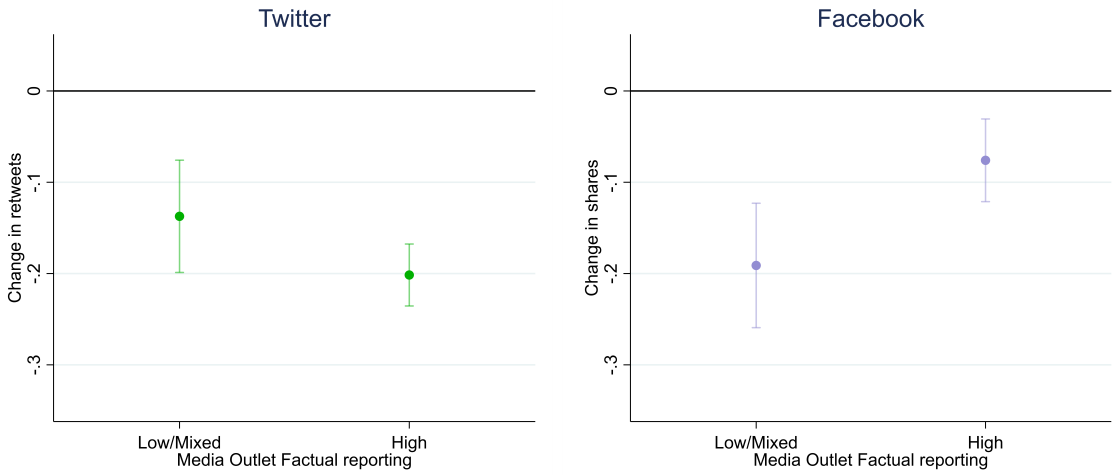
Notes: The scatter points show the results of a regression of retweets on day fixed effects (and media outlet fixed effects, not shown). In addition, kernel-weighted local polynomials fit these day-fixed-effects estimates (separately for days before, and for days after the policy changes).

low/mixed factualness outlets. We also estimate a version of this regression where we include a full set of (daily) time fixed effects. These fixed effects absorb $Post_t$ but still allow us to estimate the differential effect measured by α_1. Finally, we also estimate a version of this regression that produces two marginal effects: one for low/mixed factual outlets and one for high factual outlets.

Estimates of Equation 2, with and without the full set of time fixed effects are in Columns (2) and (3) of Table 1. Estimates of the two marginal effects for low/mixed factualness outlets and for high factualness outlets are shown in Figure 2.[11] The figure reveals clear heterogeneity in the effects by outlet type and by platform. On Twitter, high-factualness outlets experienced larger reductions in sharing relative to low/mixed factualness outlets. In particular, sharing of tweets by low/mixed factualness outlets fell by approximately 13 percent and the sharing of tweets by high factualness outlets fell by approximately 20 percent. Our estimate of α_1 is negative and statistically significant for Twitter (Table 1, Columns 2 and 3). On Facebook we find the opposite pattern. Sharing of posts by low/mixed factualness outlets fell by nearly 20 percent, but sharing of posts by high factualness outlets fell by less than 10 percent, and α_1 is positive and statistically significant at the 99% confidence level. These results suggest that Face-

[11] Similar to Figure 1, we show factualness-specific estimates of time fixed effects from a regression of retweets/shares on 59 time dummies and additional controls in the Appendix.

Figure 2: Estimated effect by media factualness



Notes: This figure shows the marginal effects of Twitter/Facebook policies for different factualness levels. It is based on a regression of retweets/shares on interactions between factualness dummies (low/mixed and high) and a dummy equal to 1 after the implementation of the policy on Twitter/Facebook. 95% confidence intervals are shown. Additional controls include time fixed effects, media outlet fixed effects and post controls, not shown.

book's policy effectively targeted content sharing for outlets that publish misleading or factually inaccurate information, but Twitter's policy did not.

We also test whether the policies' effects were heterogeneous across outlets of different political ideological lean. Political lean is an important dimension to study, as social media platforms are often criticized for creating echo chambers that amplify political divisions and partisan bias (Bail et al., 2018; Cinelli et al., 2021; Levy, 2021). This concern was particularly relevant around the 2020 US election and was explicitly addressed by Facebook who saw their role as "prevent[ing]... content, wittingly or otherwise, from aiding and abetting... violence and civil strife" (USAToday.com). An ideal policy targeting the spread of political news information should potentially affect content made by politically divisive outlets relatively more than content made by less divisive outlets.

We study this heterogeneity using the measure of outlet-level political slant $\in$ {left, lean left, center/mixed, lean right, right}. We first define a continuous variable $Slant_o$, which is monotonically increasing in political slant and varies from 0 (left) to 1 (right). We estimate the regression:

$$log(shares)_{iot} = \alpha_0 Post_t + \alpha_2 Post_t \times \text{Slant}_o + X_i'\beta + \delta_o + \epsilon_{iot} \quad (3)$$

where the key parameter $\alpha_2$captures the differential effect of a policy on a right outlet where $\text{Slant}_o = 1$ relative to a left outlet where $\text{Slant}_o = 0$.

In an alternative specification, we use a set of five dummies where each dummy captures a particular political slant, interacted with the $Post_t$ policy time dummy. This approach allows us to recover five separate parameters capturing a slant-specific change in retweets/shares after a platform's policy is implemented. Again, we estimate the regressions separately for Twitter and Facebook around their respective policy changes.

We show slant-specific marginal effects in Figure 3.[12] The two figure panels show clear differences in the effects of Twitter's and Facebook's policies. On Facebook, there were no statistically significant changes in the sharing of centrist outlets at the 95% confidence level. Most of the effects are driven by content produced by more ideologically extreme outlets. Sharing fell most for right outlets, with declines of over 20 percent, but sharing for left outlets also fell by over 10 percent. Outlets that are less ideologically extreme on both the left and the right observed smaller effects. On Twitter, the effect essentially varies monotonically with outlet slant. Content sharing fell by nearly 25 percent for left outlets, 20 percent for lean-left outlets, 15 percent for centrist or lean-right outlets and only 10 percent for right outlets.

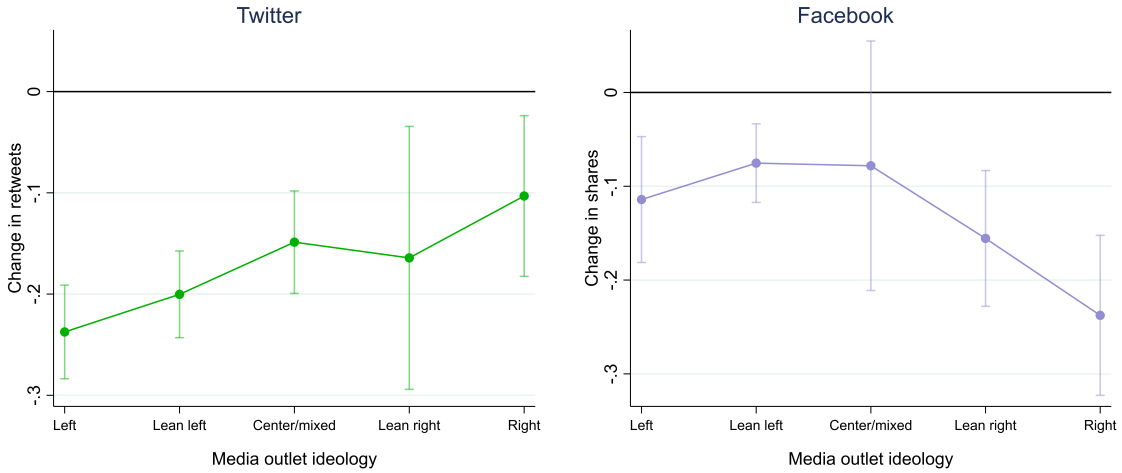
We also show estimates from the regressions with the continuous slant variable (as in Equation 3) in Columns (4), (5) and (6) of Table 1. In Columns (4) and (5) we show estimates of α_2 without and with full time fixed effects. In Column (6), we show estimates of α_2 with full time fixed effects and also a larger sample of outlets for whom we have slant information but no factualness data. Estimates are consistent across the different specifications and samples. Our estimates of α_2 are statistically significant and positive for Twitter and statistically significant and negative for Facebook. Both sets of coefficients are large in magnitude relative to the baseline α_0 in Column (4).

Outlet political slant helps explain the diverging effects of Twitter's and Facebook's policies on sharing by factualness. Figure 4 shows the share of high, medium, and low factualness outlets separated by slant. It shows that more ideologically extreme outlets have lower factualness on average. Over 90 percent of centrist outlets are classified as highly factual, compared to fewer than 20 percent of right outlets. On average, right outlets are also less factual than left outlets.

We further evaluate the heterogeneity in factualness and political slant, together, by

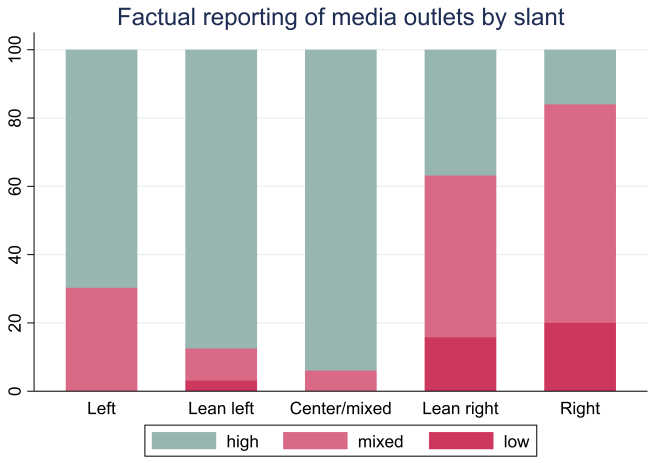
[12]Similar to Figure 1, we show estimates of slant-specific estimates of time fixed effects from a regression of retweets/shares on 59 time dummies and additional controls in the Appendix.

Figure 3: Estimated effect by media slant



Notes: This figure shows the marginal effects of Twitter/Facebook policies for different political slants. It is based on a regression of retweets/shares on interactions between slant dummies (left, lean-left, center, lean-right and right) and a dummy equal to 1 after the implementation of the policy on Twitter/Facebook. 95% confidence intervals are shown. Additional controls include time fixed effects, media outlet fixed effects and post controls, not shown.

Figure 4: Media Slant and Factualness



Notes: This figure shows the distribution of outlet factualness for each level of political slant. Relevant outlets include all outlets with MBFC factualness and Allsides political slant data. Lists of outlets by political slant and by factualness are in Tables A4 and A3.

estimating the following regression:

$$log(shares)_{iot} = \alpha_0 Post_t + \alpha_1 Post_t \times \text{High-Factualness}_o + \alpha_2 Post_t \times \text{Slant}_o + X_i'\beta + \delta_o + \epsilon_{iot} \quad (4)$$

As in Equation 2, α_1 captures the differential effect that a platform's policy has on a high-factualness outlet relative to a low/mixed-factualness outlet. Unlike the earlier regression, α_1 in Equation 4 estimates this effect conditional on the change in sharing occurring because of an outlet's political slant. Similarly, α_2 has the same interpretation as in Equation 3 but conditional on the change occurring because of an outlet's factualness. We show estimates from this regression in Column (7) of Table 1. We also show estimates from a regression that includes a full set of time dummies and absorbs the non-interacted $Post_t$ indicator in Column (8).

On Facebook, an outlet's political slant does not matter conditional on its factualness. However, more factual outlets experienced smaller reductions in their sharing, conditional on their political slant. These effects are statistically significant at the 90% confidence level and are large in magnitude: in Column (7) in the Facebook panel, the baseline α_0 is equal to -0.143, and α_1 is equal to 0.093, suggesting that high factualness outlets experienced less than half of the reduction in sharing relative to low/mixed factualness outlets.

On Twitter, we find the opposite pattern: conditional on an outlet's political slant, its factualness does not change the policy's effects. All of the heterogeneity is driven by political slant, even conditional on factualness. The coefficient on slant is statistically significant at the 99% confidence level and large in magnitude. Conditional on an outlet's change in sharing due to its factualness, the change in sharing of a right outlet's posts is less than half of a left outlet's. Together, the estimates imply that content sharing for high-factualness left outlets fell by more than for low/mixed factualness right outlets.

In the online appendix we show that our results are robust to alternative specifications, including alternative measures of factualness and political slant from Ad Fontes Media and Grinberg et al. (2019). We also document similarities in Facebook and Twitter users' sharing behaviour before either policy was implemented and show that our main results are not driven by heterogeneity in pre-policy outlet popularity. We also show that Twitter's policy reversal in December 2020 changed sharing patterns in largely the opposite direction of the effects documented here, further validating our findings.[13]

[13]To the best of our knowledge, Facebook has not reversed its policy but, unlike Twitter's, its effects

Table 1: Effect of social media platforms' policy changes on news sharing

	Twitter							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post	-0.178***	-0.137***		-0.233***			-0.210***	
	(0.018)	(0.031)		(0.019)			(0.031)	
Post x High-factual reporting		-0.064**	-0.062*				-0.025	-0.021
		(0.032)	(0.032)				(0.030)	(0.030)
Post x Slant				0.127***	0.129***	0.136***	0.111***	0.115***
				(0.042)	(0.042)	(0.033)	(0.041)	(0.042)
N	284369	284369	284369	284369	284369	508995	284369	284369
N-outlets (clusters)	137	137	137	137	137	329	137	137
R-sq	0.702	0.702	0.703	0.702	0.703	0.708	0.702	0.703
	Facebook							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post	-0.125***	-0.191***		-0.068**			-0.143***	
	(0.022)	(0.034)		(0.027)			(0.052)	
Post x High-factual reporting		0.115***	0.114***				0.093*	0.091*
		(0.042)	(0.042)				(0.049)	(0.049)
Post x Slant				-0.130**	-0.131**	-0.145***	-0.080	-0.082
				(0.053)	(0.052)	(0.044)	(0.062)	(0.062)
N	196912	196912	196912	196912	196912	329355	196912	196912
N-outlets (clusters)	136	136	136	136	136	290	136	136
R-sq	0.854	0.854	0.855	0.854	0.855	0.862	0.854	0.855
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Outlet FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day FE	No	No	Yes	No	Yes	Yes	No	Yes
Sample	Main	Main	Main	Main	Main	Extnd.	Main	Main

Notes: OLS regressions with number of retweets/shares (log) as the outcome. *Post* is a dummy indicator equal to 1 after the implementation of the policy on Twitter/Facebook. High-factualness is a dummy equal to 1 if the outlet has "high" factualness score. Slant is defined as a continuous variable which varies from 0 (left) to 1 (right). Controls include number of likes/engagements (log), whether the tweet/post contains a url link, a hashtag, an @ mention, whether the tweet is a reply or a retweet, and tweet length. Standard errors clustered at the media outlet level. Significance levels indicated *p<0.10, ** p<0.05, ***p<0.01.

5 Discussion

Our paper documents the effects of two different social media platforms' policies aiming to mitigate the spread of misinformation. Both policies reduced news media content sharing, but Facebook's "top-down" policy more successfully targeted content made by less factual outlets and more politically extreme outlets. Twitter's policy was less successful, suggesting that users who follow right/right-wing outlets (and who are likely more conservative) were less responsive to the change in the UI, relative to those following more left outlets. The question remains: why did more conservative users fail to respond to Twitter's policy?

Our outlet-focused data does not allow us to provide a clear answer to this question but substantial previous work shows heterogeneity in social media usage by political affiliation. Conservatives are more likely both to be exposed to and to share false or misleading information (Grinberg et al., 2019; Guess, Nagler and Tucker, 2019; Guess, Nyhan and Reifler, 2020; Henry ⓡ, Zhuravskaya ⓡ and Guriev, 2020), while they report higher levels of social media usage (Conzo ⓡ et al., 2021). Partisanship has also been shown to be an important correlate of fake news sharing (Osmundsen et al., Forthcoming). Experimental studies point to similar explanations. Trump supporters were less trustful of fact-checking and less responsive to a fact-checking intervention (irrespective of news congeniality, Clayton et al., 2020). Similarly, an intervention which shifted participants' attention to accuracy generally reduced participants' sharing intentions of false headlines; however, this treatment did not significantly reduce sharing intentions of Republicans exposed to concordant news (Pennycook et al., 2021). Arguably, these are precisely the Twitter users who continued to retweet right-wing news media despite Twitter's prompts, and it is also consistent with the idea that conservative partisan outlets do not tend to compete on the dimension of accuracy (Faris et al., 2017; Osmundsen et al., Forthcoming).

Conservatives are also more likely to mistrust the actions of social media platforms. According to a 2020 Pew survey, 70 percent of self-identified Republicans believe that "major tech companies support the views of liberals over conservatives," and nearly 90 percent of Republicans believe that "social media sites likely censor political views" and have no confidence in the ability of "social media companies to determine which posts on their platforms should be labeled as inaccurate or misleading" (PewResearch.org).

were limited to political news. In the Appendix, we show that the policy had no effect on sharing patterns of non-political posts made by a sample of NFL football players.

This perception was held and amplified by many prominent conservative politicians.[14] Conservative politicians and publications also responded strongly to Twitter's October 2020 policy.[15] These differences suggest, once again, that more conservative users may have been less likely to follow Twitter's encouragement of "more consideration" before sharing content on the platform.

6 Conclusion

Our findings have substantial implications for policies that attempt to mitigate online political polarization and the spread of misinformation. If platforms (or governments wishing to regulate platforms) want to implement policies that rely on user actions, they need to carefully consider the potential heterogeneity in user responses along relevant dimensions. In our study, more conservative Twitter users appeared less likely to follow the platform's prompts and continued to share right-wing news, leading to both an imbalance of the political effects of the policy and adverse overall effects on the sharing of misleading news relative to more factual news. These results highlight the potential unintended consequences of platforms' interventions.

Facebook's policy did not rely on active user participation and appeared to effectively restrict the sharing of content produced by outlets with less factual reporting, and of content produced by outlets with more biased political slant, potentially resulting in socially beneficial outcomes. But these results also highlight how powerful such policy interventions can be, and the ability of platforms to affect content spread through algorithmic adjustments. Subtle (and invisible to the user) changes in a platform's algorithms can affect the reach of content without actively interfering with users' behaviour.[16] More broadly, the popularity of social media posts may affect individuals' policy preferences (Conzo ⓡ et al., 2021), news consumption (Messing and Westwood, 2014), online political expression (Morales, 2020) and traditional media coverage (Cagé,

[14] In 2020 Ted Cruz "has held hearings on allegations that social media companies 'censor' conservative speech online" (Politico.com). In May 2020, Trump signed an executive order that aimed "to limit the companies' legal immunity for how they moderate content posted by users," stating that "it's been very unfair." (LATimes.com).

[15] Republican congressman Doug Collins and the official account of the Judiciary Committee Republicans both claimed that Twitter was censoring articles from conservative political commentator Sean Hannity's website, Hannity.com (TheVerge.com). Other prominent conservatives had similar responses (Reuters.com). To the best of our knowledge, despite some confusion about Twitter's policy there was no comparable response among Democratic politicians or commentators (Slate.com).

[16] For example, TikTok has been accused of suppressing content related to Hong Kong protests or the Chinese government's treatment of Uighurs (WashingtonPost.com). TikTok's response was that such content was simply not going viral on their platform.

Hervé and Mazoyer, 2020). Our findings raise concerns about the power of social media platforms to steer public opinion, political dynamics and media decisions.

References

Allcott, Hunt, and Matthew Gentzkow. 2017. "Social media and fake news in the 2016 election." *Journal of Economic Perspectives*, 31(2): 211–36.

Bail, Christopher A, Lisa P Argyle, Taylor W Brown, John P Bumpus, Haohan Chen, MB Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, and Alexander Volfovsky. 2018. "Exposure to opposing views on social media can increase political polarization." *Proceedings of the National Academy of Sciences*, 115(37): 9216–9221.

Baly, Ramy, Georgi Karadzhov, Dimitar Alexandrov, James Glass, and Preslav Nakov. 2018. "Predicting factuality of reporting and bias of news media sources." *arXiv preprint arXiv:1810.01765*.

Baly, Ramy, Georgi Karadzhov, Jisun An, Haewoon Kwak, Yoan Dinkov, Ahmed Ali, James Glass, and Preslav Nakov. 2020. "What was written vs. who read it: News media profiling using text analysis and social media context." *arXiv preprint arXiv:2005.04518*.

Cagé, Julia, Nicolas Hervé, and Béatrice Mazoyer. 2020. "Social Media and Newsroom Production Decisions." *Available at SSRN 3663899*.

Cagé, Julia, Nicolas Hervé, and Marie-Luce Viaud. 2020. "The Production of Information in an Online World." *The Review of Economic Studies*, 87(5): 2126–2164.

Cinelli, Matteo, Gianmarco De Francisci Morales, Alessandro Galeazzi, Walter Quattrociocchi, and Michele Starnini. 2021. "The echo chamber effect on social media." *Proceedings of the National Academy of Sciences*, 118(9).

Clayton, Katherine, Spencer Blair, Jonathan A Busam, Samuel Forstner, John Glance, Guy Green, Anna Kawata, Akhila Kovvuri, Jonathan Martin, Evan Morgan, et al. 2020. "Real solutions for fake news? Measuring the effectiveness of general warnings and fact-check tags in reducing belief in false stories on social media." *Political Behavior*, 42(4): 1073–1095.

Conzo ⓡ, Pierluigi, Laura K Taylor ⓡ, Juan S Morales ⓡ, Margaret Samahita ⓡ, and Andrea Gallice. 2021. "Can ❤s Change Minds? Social Media Endorsements and Policy Preferences." *Carlo Alberto Notebooks n. 641*.

Faris, Robert, Hal Roberts, Bruce Etling, Nikki Bourassa, Ethan Zuckerman, and Yochai Benkler. 2017. "Partisanship, propaganda, and disinformation: Online media and the 2016 US presidential election." *Berkman Klein Center Research Publication*, 6.

Grinberg, Nir, Kenneth Joseph, Lisa Friedland, Briony Swire-Thompson, and David Lazer. 2019. "Fake news on Twitter during the 2016 US presidential election." *Science*, 363(6425): 374–378.

Guess, Andrew, Jonathan Nagler, and Joshua Tucker. 2019. "Less than you think: Prevalence and predictors of fake news dissemination on Facebook." *Science Advances*, 5(1).

Guess, Andrew M, Brendan Nyhan, and Jason Reifler. 2020. "Exposure to untrustworthy websites in the 2016 US election." *Nature Human Behavior*, 4(5): 472–480.

Henry ⓡ, Emeric, Ekaterina Zhuravskaya ⓡ, and Sergei Guriev. 2020. "Checking and sharing alt-facts." *Available at SSRN*.

Levy, Ro'ee. 2021. "Social Media, News Consumption, and Polarization: Evidence from a Field Experiment." *American Economic Review*, 111(3): 831–70.

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

Morales, Juan S. 2020. "Perceived Popularity and Online Political Dissent: Evidence from Twitter in Venezuela." *The International Journal of Press/Politics*, 25(1): 5–27.

Morales, Juan S. 2021. "Legislating during war: Conflict and politics in Colombia." *Journal of Public Economics*, 193: 104325.

Osmundsen, Mathias, Alexander Bor, Peter Bjerregaard Vahlstrup, Anja Bechmann, and Michael Bang Petersen. Forthcoming. "Partisan polarization is the primary psychological motivation behind "fake news" sharing on Twitter." *American Political Science Review*.

Pennycook, Gordon, Ziv Epstein, Mohsen Mosleh, Antonio A Arechar, Dean Eckles, and David G Rand. 2021. "Shifting attention to accuracy can reduce misinformation online." *Nature*, 1–6.

Vosoughi, Soroush, Deb Roy, and Sinan Aral. 2018. "The spread of true and false news online." *Science*, 359(6380): 1146–1151.

Zhuravskaya, Ekaterina, Maria Petrova, and Ruben Enikolopov. 2020. "Political effects of the internet and social media." *Annual Review of Economics*, 12: 415–438.

A Appendix (For Online Publication)

A.1 Description of factualness and political slant data sources

Our data on outlet-level political bias/slant comes from Allsides. Allsides uses a combination of methods to give a "bias" rating to each outlet (Allsides.com). They conduct blind surveys where respondents rate their own political bias and the political bias of articles by unknown media outlets, and then average out bias ratings across media outlets. In addition to this, they occasionally use editorial review by their own editors, and by other external sources such as Pew. They also incorporate community feedback: visitors to Allsides.com can indicate whether they agree or disagree with an outlet's rating. We restrict the sample to the most popular outlets, with a ratio of *bias* agreement of at least 50 percent, that is, most users agreed with the bias rating given by AllSides. The list of media outlets and their political slant ratings is shown in Table A3 in the Appendix.

Our data on factualness comes from Media Bias Fact Check (MBFC). They use editorial fact-checking of outlets' articles to rate the factualness of outlets. According to MBFC, a high-factualness outlet uses "factual sources," makes "immediate corrections to incorrect information" and has failed at most one fact check on a sample of at least 5 news stories (MediaBiasFactCheck.com). A mixed-factualness outlet "does not always use proper sourcing" or combines credible and non-credible information, has failed more than one fact check on a sample of at least 5 news stories and does not always correct factual mistakes. MBFC also labels every outlet that does not disclose a mission or ownership information as mixed-factualness. A low-factualness outlet rarely or never uses "credible" sources and publishes news stories that are inaccurate, conspiracy theories and propaganda (MediaBiasFactCheck.com). In some cases, MBFC also separates high and low factualness into four categories: "very high," "high," "low" and "very low." Our data sources (Baly et al. 2018, 2020) group the two "high" and two "low" categories together. A list of media outlets and their factualness ratings are in Table A4.

In addition to MBFC, we collected data from Ad Fontes Media, another data provider that ranks both the factualness and political slant of news outlets. They use a team of hired analysts to rate articles from each outlet, and article scores were then aggregated up to outlet scores. Factualness ratings are based on a combination of "veracity" (whether the content is true, easily provable and widely accepted), and "expression" (whether the content is presented as fact, fact with some analysis, or opinion). Outlets that present untrue content as fact receive a low factualness score. Political position is

based on the language/terminology the articles use and whether they present opposing political opinions as a point of comparison. In the rating version we use, the analysts were rating independently and they were selected to be representative of national left-right political opinions (adfontesmedia.com).

Ad Fontes has a continuous measure for outlet factualness ranging between 0 for very low factualness/reliability outlets and 64 for very high reliability. To be comparable to the MBFC classification, we discretize this measure into two bins. We label outlets with a score higher than 32 as "high factualness" and outlets with a score lower than 32 as "low factualness." Ad Fontes also has a continuous measure of outlet political slant, which ranges from -42 for the most left outlets to 42 for the most right outlets. To be comparable to the Allsides classification, and following Ad Fontes' own classifications (see adfontesmedia.com) we discretize this measure into five bins. We label outlets with scores less than -16.5 as "left," outlets with scores between -16.5 and -5.5 as "lean-left." We classify outlets with scores between -5.5 and 5.5 as "centrist," and outlets with scores between 5.5 and 16.5, and outlets with scores above 16.5 as "lean-right" and "right," respectively.

We also collected data on outlet factualness from Grinberg et al. (2019). As part of their analysis, they classify a large number of media outlets as "fake news" outlets and "non-fake news" outlets. Fake news outlets are defined as outlets that "were likely to share political misinformation... due to poor journalistic practices" (Grinberg et al. 2019 Appendix S.5).[17] This classification was based on pre-existing lists of outlets from other academic papers and fact checking organizations, as well as additional classification done by the authors based on snopes.com fact checking.

Ad Fontes ranks fewer outlets than MBFC and Allsides, so we do not use it in our main estimates but we present results with their factualness and slant ratings as part of our robustness checks. Grinberg et al. (2019)'s coverage is closer to MBFC, but the goal of their classification is narrower than MBFC's or Ad Fontes'. Grinberg et al. (2019) is meant to identify fake news producers, rather than generally evaluate media outlets' quality. Under this restrictive classification a very large majority of the popular Allsides outlets are classified as non-fake, including a number of outlets classified by Ad Fontes and MBFC to be unreliable or low factualness (e.g., Newsmax and OANN). Nonetheless, capturing the effects of platform policies on content made by fake news websites is still

[17] In the paper, there are several classes of fake news outlets which capture the degree of misinformation produced by the outlet: yellow, orange, red and black. Non-fake news outlets are classified as green. Since there is a very small number of Grinberg et al. (2019) fake news outlets in our data, we aggregate them into one category.

important and we present results with the Grinberg et al. (2019) factualness classification as part of our robustness checks. MBFC also constructs its own political bias/slant measure for outlets. Since there are fewer outlets in the MBFC data than in our main Allsides data, we choose to use Allsides data as the main slant measure but we also present results with the MBFC slant ranking as part of our robustness checks. The estimates using alternative sources for factualness and slant are consistently similar with our main estimates.

A.2 Heterogeneity by outlet baseline engagement

There are a number of concerns related to both the heterogeneity of effects found *within* a given platform, and the heterogeneity of effects found across platforms' policies. One concern is reversion to the mean: if "left" outlet content is *more* popular and shared *more* on Twitter than "right" outlet content, then we cannot find a big effect for "right" outlets because they are not shared enough. If on Facebook the situation is reversed and "right" outlet content is more popular/shared than "left" outlet content, then we will mechanically find a bigger effect of Facebook's policy on "right" outlets. A second concern is our choice of looking at relative rather than absolute changes in sharing by using $log(shares + 1)$ as the outcome variable: Looking at relative changes could potentially lead to "mechanically" heterogeneous findings as well. These would go in the opposite direction of the reversion to the mean argument. If "left" outlet content is *less* popular and shared *less* on Twitter than "right" outlet content, then similar absolute changes in the number of retweets would produce a bigger relative effect on "left" outlets than on "right" outlets. A third concern is about user heterogeneity between Facebook and Twitter: Facebook has more users than Twitter, and Twitter's users may be a more selected sample with different characteristics. In particular, Twitter's users could be more ideologically committed, which would directly affect the results.

We first examine possible heterogeneity between Facebook and Twitter users by looking at sharing behaviour in the period before both platforms' policies were enacted. We focus on the first week of October 2020, which also predates Twitter's announcement of its policy change on October 9 (blog.twitter.com). We refer to this period as the "baseline" period. Figure A10 compares outlet-level average sharing on Facebook and on Twitter during the baseline period and shows very strong correlation in media sharing between the two platforms. On average, media outlets with posts that are more shared on Facebook in early October are also the outlets with posts that are more shared on Twitter and the correlation coefficient between the two is approximately 0.7.

This figure also shows that on both platforms, content made by low/mixed-factualness outlets is shared more than content made by high-factualness outlets. This suggests that despite differences in platform size and user characteristics, the two user-bases display similar behaviour with respect to our main variable of interest. This also suggests that differences in estimated effects between Facebook and Twitter were not driven by differences in ex-ante sharing behaviour.

Next, we evaluate concerns that the heterogeneity in effects we find within each platform is primarily caused by reversion to the mean, or by examining *relative* rather than *absolute* changes in sharing. Both of these concerns relate to heterogeneity in pre-policy outlet sharing and we address them by estimating a new set of regressions that allow for the effects of the policy to vary by an outlet's popularity in the "baseline" period. In addition to $(Post_t \times \text{Slant}_o)$, $(Post_t \times \text{High-Factualness}_o)$, we introduce the interaction $(Post_t \times \text{Baseline Shares}_o)$ which captures an outlet's mean number of shares in the first week of October 2020.[18]

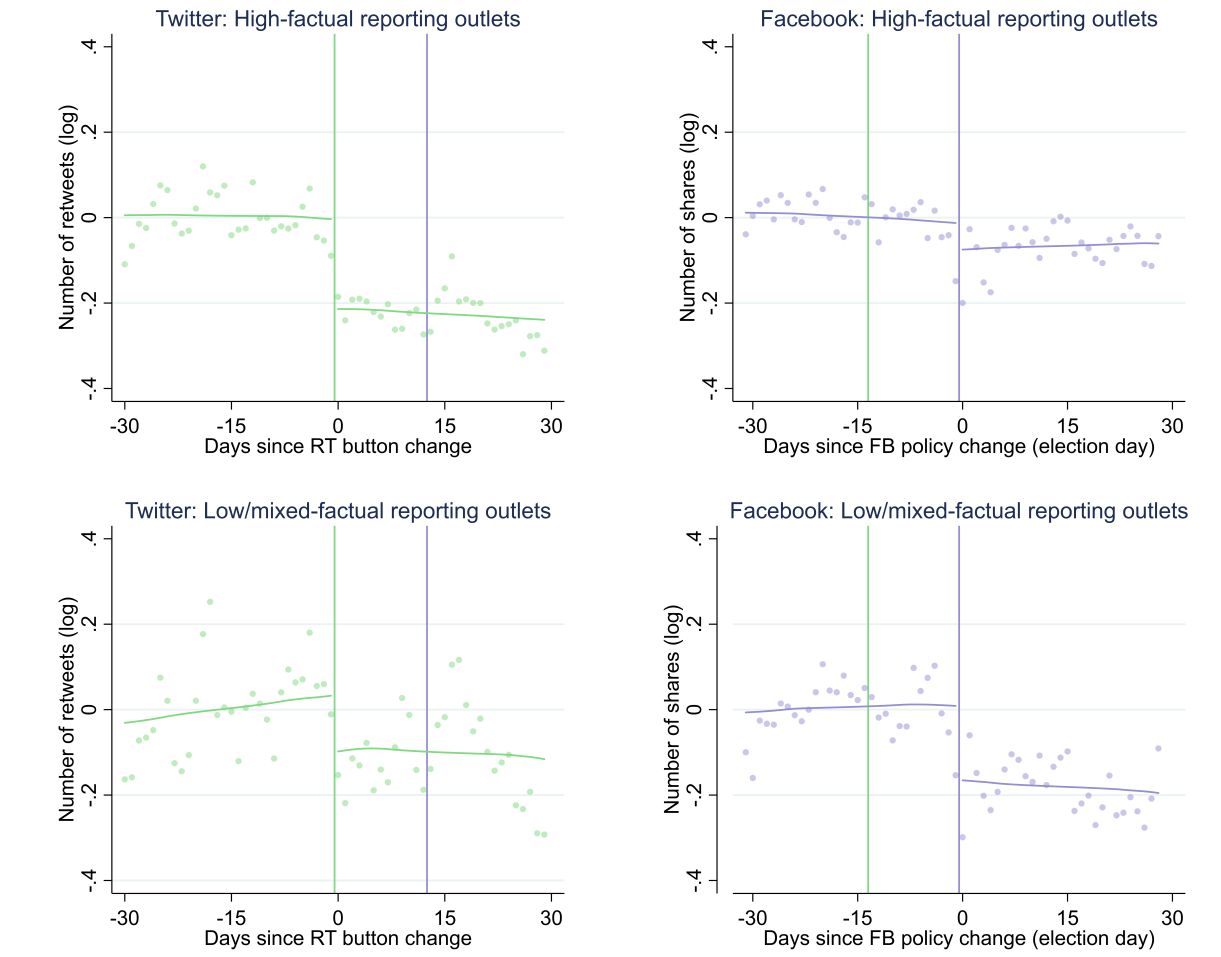
Estimates from these regressions are in Tables A15 and A16. For Twitter, we find that allowing the effects of the policy to vary by baseline sharing does not change our estimated coefficients for slant or for factualness relative to estimates in the main text. The coefficients on the interaction with baseline shares is statistically insignificant and small.

For Facebook, we find that point estimates for the main coefficients do not change substantively relative to the main text, but the additional interaction with baseline-sharing increases standard errors on all coefficients. However, since there are very few mixed/low-factualness outlets with low baseline-sharing, and many mixed/low-factualness outlets with very high baseline sharing (see Figure A10), the interaction with baseline sharing absorbs much of the variation in the data. In Columns (3)-(5) of Table A16 we address this issue by restricting our sample to outlets with above-median sharing in the baseline period and find that the factualness effects are similar to those in the main text, and that the interaction with baseline engagement becomes statistically insignificant. We also show that the finding that Facebook's policy was effective at reducing low/mixed factualness outlet sharing relatively more is also robust to using the Ad Fontes alternative measure/sample for factualness (Column 2).

[18] Due to the collection limits imposed by the Twitter API (the last 3,200 tweets) our sample becomes slightly smaller, 111 outlets.

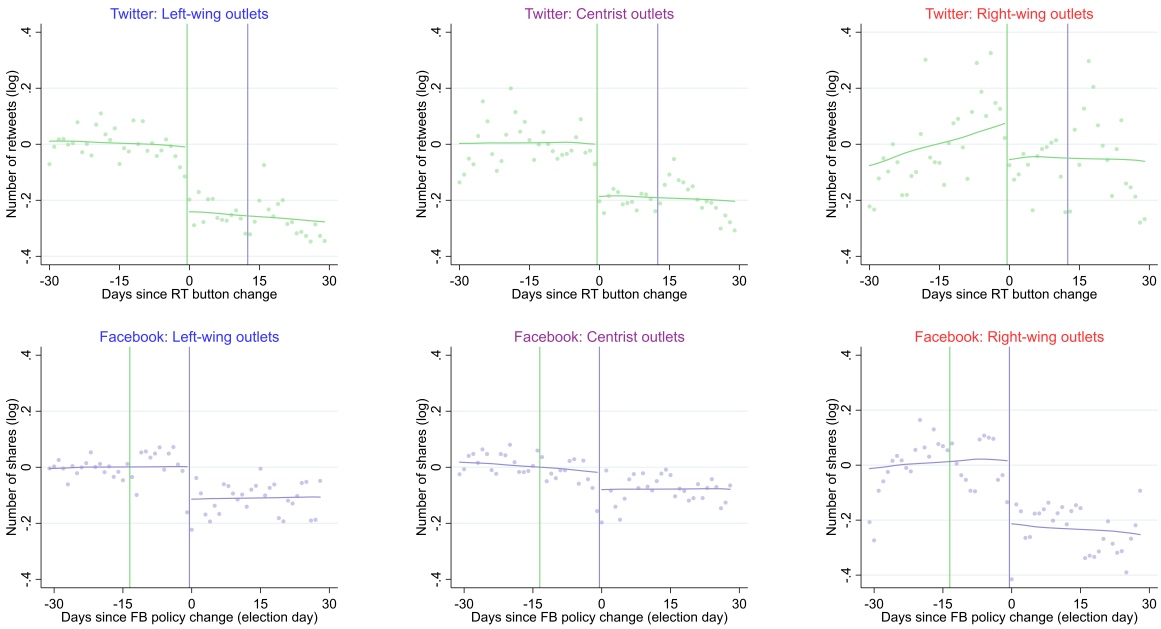
A.3 Additional figures and tables

Figure A1: Social media platforms' policy changes and news sharing by factualness



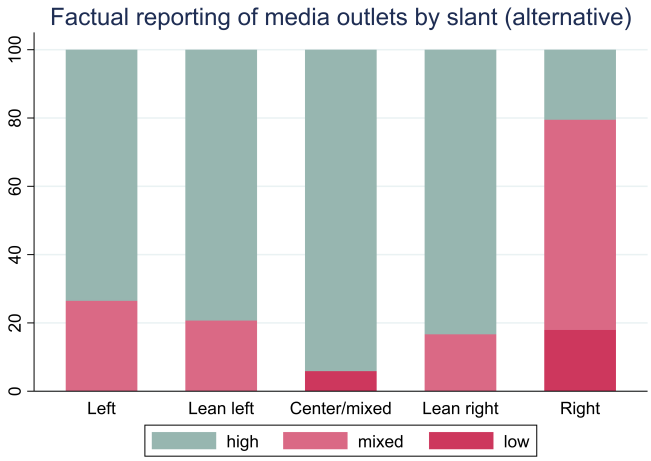
Notes: Each panel uses a factualness specific sample (i.e., only high-factualness outlets) and plots the results of a regression of retweets on day fixed effects (and media outlet fixed effects, not shown). In addition, kernel-weighted local polynomials fit these day-fixed-effects estimates (separately for days before, and for days after the interface change).

Figure A2: Policy changes and news sharing by media outlet slant



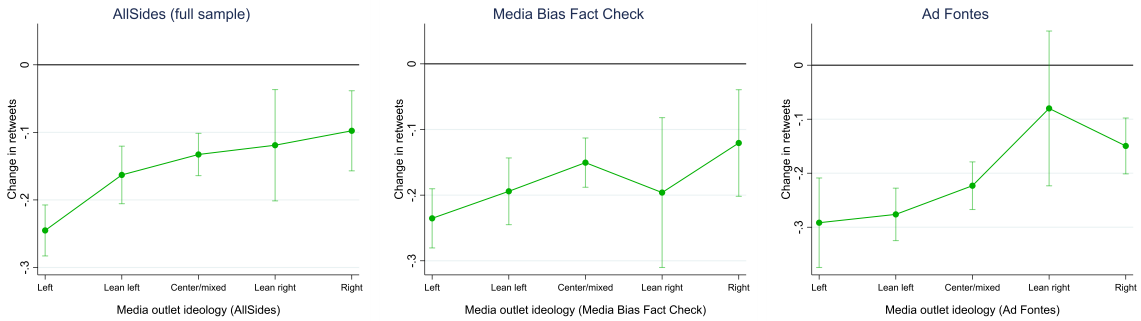
Notes: Each panel uses a political-slant specific sample (i.e., only left-wing outlets) and plots the results of a regression of retweets on day fixed effects (and media outlet fixed effects, not shown). In addition, kernel-weighted local polynomials fit these day-fixed-effects estimates (separately for days before, and for days after the interface change).

Figure A3: Media Slant and Factualness (Media Bias Fact Check)



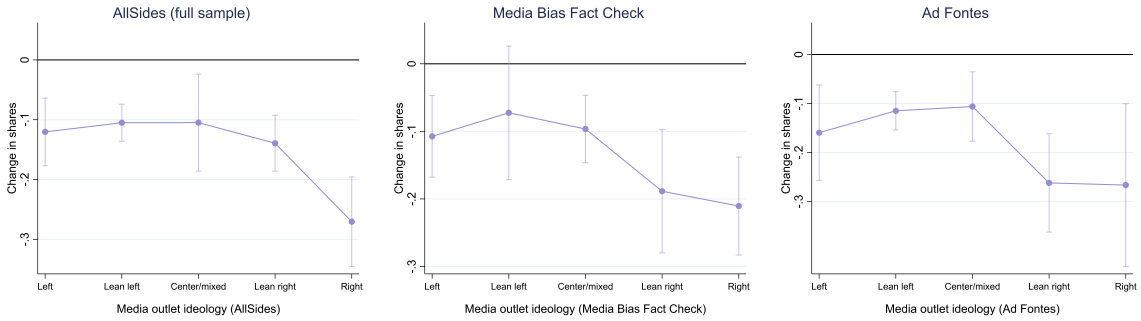
Notes: This figure shows the distribution of outlet factualness for each level of political slant. Relevant outlets include all outlets with MBFC factualness and political slant data.

Figure A4: Estimated effect of Twitter UI change by media slant (alternative measures)



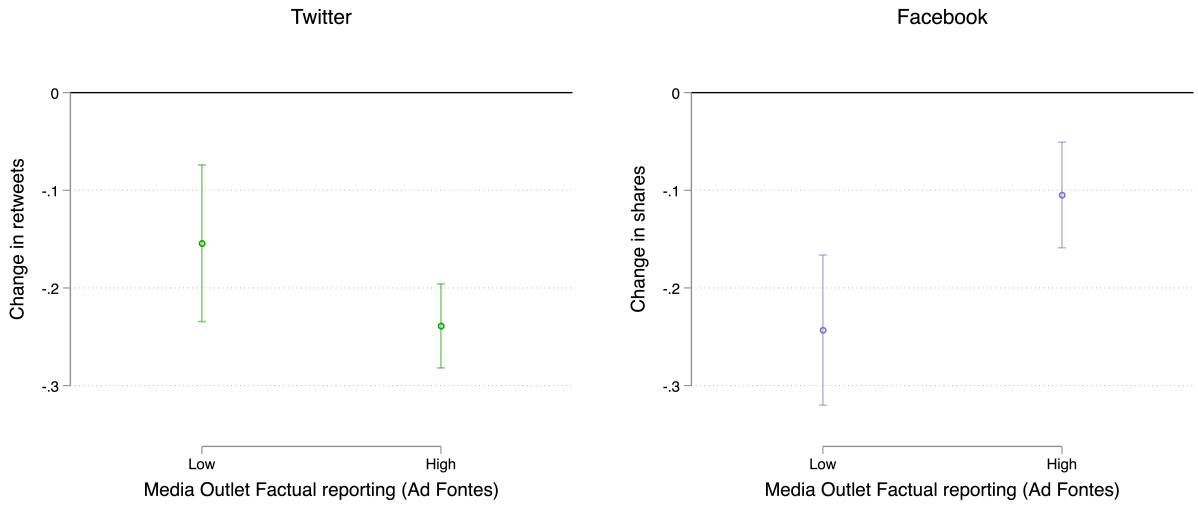
Notes: This figure shows the marginal effects of Twitter's policy intervention for different political slants. Each of the panels uses a different data source for outlet-level political slant: AllSides, Media Bias Fact Check and Ad Fontes. Each set of results is based on a regression of retweets/shares on interactions between slant dummies (left, lean-left, center, lean-right and right) and a dummy equal to 1 after the implementation of a policy on Twitter. 95% confidence intervals are shown. Additional controls include time fixed effects, media outlet fixed effects and post controls, not shown.

Figure A5: Estimated change in shares after FB policy intervention (election day) (alternative measures)



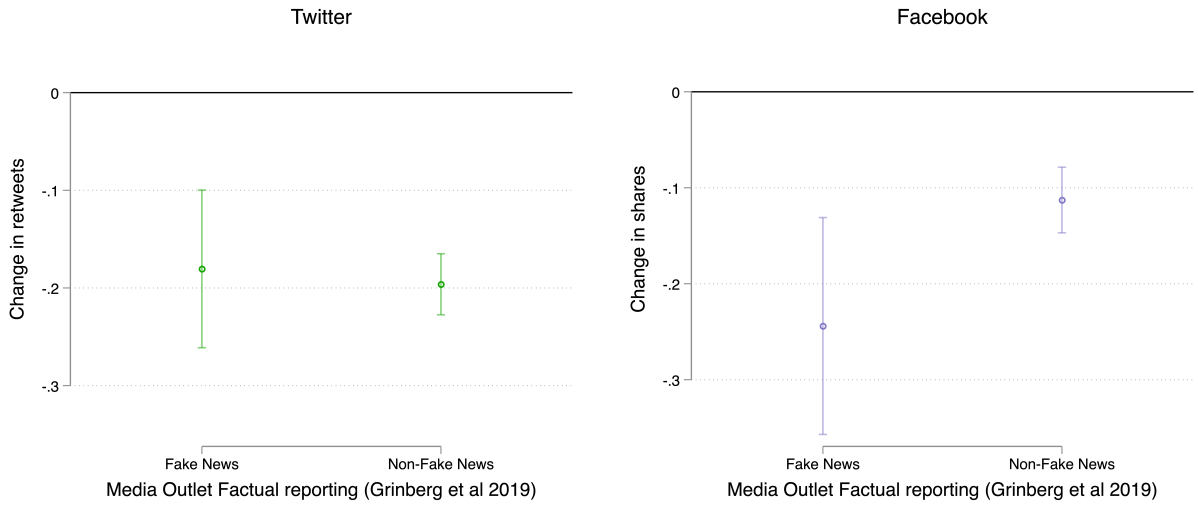
Notes: This figure shows the marginal effects of Facebook's policy intervention for different political slants. Each panel uses a different data source for outlet-level political slant: AllSides, Media Bias Fact Check and Ad Fontes. Each set of results is based on a regression of retweets/shares on interactions between slant dummies (left, lean-left, center, lean-right and right) and a dummy equal to 1 after the implementation of a policy on Facebook. 95% confidence intervals are shown. Additional controls include time fixed effects, media outlet fixed effects and post controls, not shown.

Figure A6: Estimated change in shares after policy intervention (Ad Fontes)



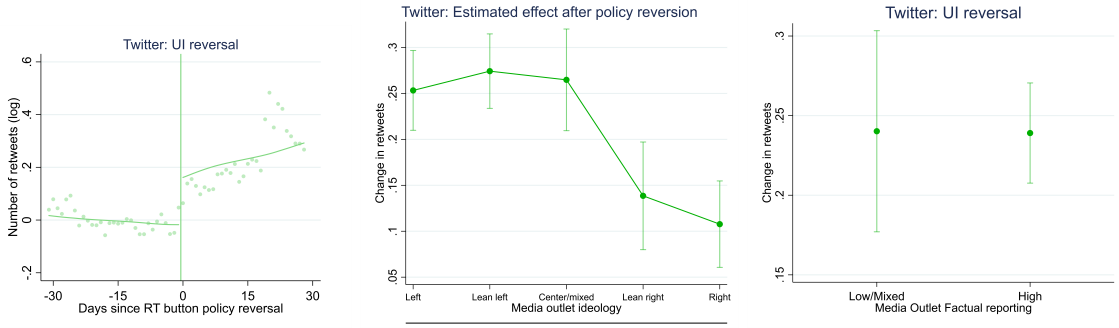
Notes: This figure shows the marginal effects of Twitter/Facebook policies for different factualness levels. Unlike in Figure 2 in the main text, the data source for factualness is Ad Fontes. Estimates are based on a regression of retweets/shares on interactions between factualness dummies (low and high) and a dummy equal to 1 after the implementation of the policy on Twitter/Facebook. A low factualness outlet has a below mean Ad Fontes factualness score. 95% confidence intervals are shown. Additional controls include time fixed effects, media outlet fixed effects and post controls, not shown.

Figure A7: Estimated change in shares after policy intervention (Grinberg et al 2019)



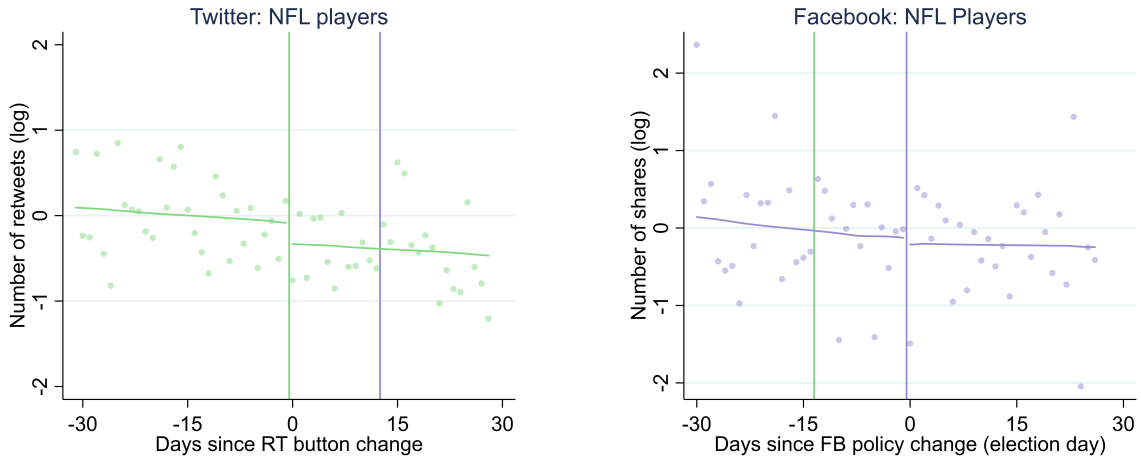
Notes: This figure shows the marginal effects of Twitter/Facebook policies for different factualness levels. Unlike in Figure 2 in the main text, the data source for factualness is Grinberg et al. (2019). Estimates are based on a regression of retweets/shares on interactions between factualness dummies and a dummy equal to 1 after the implementation of the policy on Twitter/Facebook. Grinberg et al. (2019) classifies media outlets into "Fake News" and "Non-Fake News" outlets. 95% confidence intervals are shown. Additional controls include time fixed effects, media outlet fixed effects and post controls, not shown.

Figure A8: Estimated effects of Twitter UI reversal



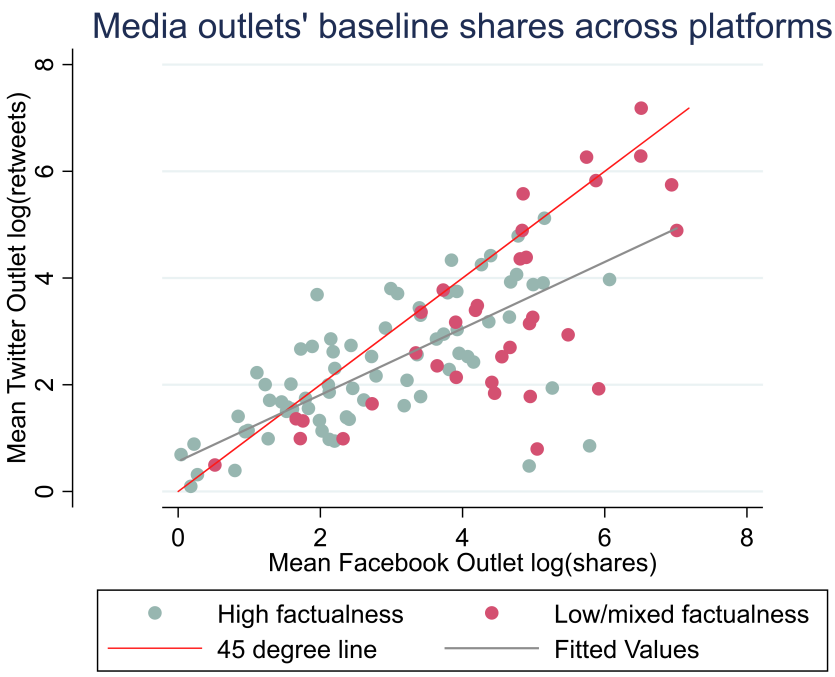
Notes: The left panel's scatter points show the results of a regression of retweets on day fixed effects in the 60 days around December 15, Twitter's policy reversal date (and media outlet fixed effects, not shown). In addition, kernel-weighted local polynomials fit these day-fixed-effects estimates (separately for days before, and for days after the interface reversal). The middle panel shows the marginal effects of Twitter policy reversal for different factualness levels. It is based on a regression of retweets/shares on interactions between factualness dummies (low/mixed and high) and a dummy equal to 1 after Twitter's policy reversal. The right panel shows the marginal effects of Twitter policy reversal for different slant levels. It is based on a regression of retweets/shares on interactions between slant dummies (left, lean-left, center, lean-right and right) and a dummy equal to 1 after Twitter's policy reversal. For the middle and right panels, 95% confidence intervals are shown and additional controls include time fixed effects, media outlet fixed effects and post controls, not shown.

Figure A9: Social media platforms' policy changes and content sharing: NFL players



Notes: The scatter points show the results of a regression of retweets/shares on day fixed effects (and media outlet fixed effects, not shown). In addition, kernel-weighted local polynomials fit these day-fixed-effects estimates (separately for days before, and for days after the interface change). The estimated changes in sharing are $\alpha = -0.419, p < 0.01$ for Twitter, and $\alpha = -0.121, p > 0.10$ for Facebook (with standard errors clustered at the player level).

Figure A10: Facebook and Twitter October 2020 Sharing Comparison



Notes: Each point in this scatter plot represents an outlet during the first week of October 2020. The horizontal axis shows the mean number of Facebook shares an outlet's posts received in the first week of October 2020 (in logs). The vertical axis shows the mean number of Twitter retweets an outlet's posts received in the first week of October 2020 (in logs).

Table A1: Twitter Summary Statistics

Variable	Obs	Mean	Std. Dev.
Post Made After Policy Change	284369	.651	.477
N Post Shares (retweets)	284369	200.547	2321.166
Outlet Slant (Left)	284369	.255	.436
Outlet Slant (Lean left)	284369	.243	.429
Outlet Slant (Center/mixed)	284369	.211	.408
Outlet Slant (Lean right)	284369	.143	.35
Outlet Slant (Right)	284369	.149	.356
High-factual Reporting Outlet	284369	.641	.48
N Likes	284369	303.162	1977.177
Post Length (chars)	284369	123.313	25.775
Post Contains Link	284369	.889	.314
Post Contains @	284369	.238	.426
Post Contains Hashtag	284369	.05	.218
Post is Retweet	284369	.788	.409

Table A2: Facebook Summary Statistics

Variable	Obs	Mean	Std. Dev.
Post Made After Policy Change	196913	.494	.5
N Post Shares	196913	331.957	2289.185
Outlet Slant (Left)	196913	.236	.425
Outlet Slant (Lean left)	196913	.265	.441
Outlet Slant (Center/mixed)	196913	.204	.403
Outlet Slant (Lean right)	196913	.09	.287
Outlet Slant (Right)	196913	.204	.403
High-factual Reporting Outlet	196913	.571	.495
N Total Engagement	196913	3110.768	13297.8
Post Length (chars)	196913	153.473	127.204
Post Contains Link	196913	.103	.303
Post Contains @	196913	.005	.072
Post Contains Hashtag	196913	.03	.171

Table A3: AllSides media outlets by Twitter handle and slant

Left	Lean left	Center/mixed	Lean right	Right
ADAction	ABC	AARP	AEI	AccuracyInMedia
ajplus	ACLU	Abridge_News	amconmag	ACUConservative
AlterNet	agerney	AllSidesNow	bostonherald	AlCardenasFL_DC
ArkansasOnline	ajc	AP	CatoInstitute	AllysiaFinley
bluevirginia	AJEnglish	axios	CSBA_	AmericanThinker
BoingBoing	amnesty	ballotpedia	Daily_Press	amspectator
BuzzFeedNews	amprog	BarnPat	dcexaminer	AnnCoulter
Care2	AndrewYang	BBCNews	DeseretNews	AtlasNetwork
CenterOnBudget	AnnafiWahed	billybinion	DickMorrisTweet	bearingdrift
ChildDefender	AP_Politics	BrookingsInst	DouthatNYT	benshapiro
CNNOpinion	bgdailynews	businessinsider	drudgefeed	BreitbartNews
curaffairs	BostonGlobe	CalMatters	EpochTimes	BrentBozell
dailykos	business	CalWatchdog	ExaminerOnline	CBNNews
democracynow	bustle	CarnegieEndow	feeonline	charliekirk11
dhnews	camanpour	chicagotribune	FoxNews	CityJournal
EconomicPolicy	CBSNews	CivilBeat	FRCdc	cnsnews
EJDionne	CentreView	CNBC	Heritage	CollegeFix
esquire	CharlesMBlow	CNET	IBDinvestors	DailyMailUK
Eugene_Robinson	CJEducation	ConstitutionCtr	IndependentInst	DailySignal
ezraklein	CNN	CookPolitical	JoeNBC	debrajsaunders
fcnp	daily_targum	countertweeter	JohnStossel	DennisPrager
HuffPost	DamonLinker	Crowdpac	JudicialWatch	DRUDGE
jacobinmag	DLeonhardt	csmonitor	kathleenparker	EdRogersDC
Jezebel	edshow	cspanradio	leesburgtoday	FDRLST
jonathanchait	EnvDefenseFund	dailycardinal	LiveActionNews	foxnewslatino
mashable	FAIRmediawatch	DailyProgress	ManhattanInst	FoxNewsOpinion
mmfa	FAScientists	dallasnews	MarketWatch	fpmag
MotherJones	Gizmodo	DefenseOne	MJGerson	frankminiter
MSNBC	googlenews	diplocourier	MrAndyNgo	FreeBeacon
NCPSSM	grist	DukeChronicle	newsmax	gatewaypundit
newrepublic	guardian	EPTrailGazette	nypost	GeorgeWill
newsone	HealthCareGov	ErikWemple	nytdavidbrooks	glennbeck
Newsweek	Independent	erumors	OpinionWSJ	GOP
NewYorker	indyweek	EurekAlert	PeterRoff	GroverNorquist
NickKristof	JRubinBlogger	FaceFactsUSA	PittsburghPG	Judgenap
NYDailyNews	LasVegasSun	FareedZakaria	Project_Veritas	KimStrassel
NYMag	latimes	FinancialTimes	Quillette	KSLcom
nytopinion	Mediaite	FiveThirtyEight	RandPaul	marcthiessen
paulkrugman	MiamiHerald	fixitshow	reason	mgoodwin_nypost
PeacockPanache	michigandaily	Forbes	sullydish	michellemalkin
peoplefor	monthly	ForeignAffairs	Telegraph	newsbusters
PNHP	NAACP	Freakonomics	TheBabylonBee	newsmax
politicususa	NBCNews	fulcrum_us	thedispatch	newtgingrich
RawStory	nytimes	GallupNews	TheFiscalTimes	NRO
RevJJackson	online_HBS	H_R_Messenger	TheIJR	OANN
RollingStone	PacificStand	HowardKurtz	TheLibRepublic	Peggynoonannyc
ryanccooper	piersmorgan	IBTimes	TPostMillennial	PJMedia_com
Salon	politico	idsnews	WashTimes	prageru
sfchronicle	PolitiFact	InsidePhilanthr	WhiteHouse	RameshPonnuru
Slate	propublica	ivn		realDailyWire
SocialistAlt	publicintegrity	KnowTheFlipSide		RealRLimbaugh
socialistprojct	RANDCorporation	KQED		RedState
splinter_news	RuthMarcus	lifehacker		RichLowry
StephenAtHome	sacbee_news	ListenFirstProj		rightsidenews
thedailybeast	sciam	maureendowd		sallypipes
TheDailyShow	SFGate	Milbank		seanhannity
theintercept	ShowUngar	nationaljournal		SpeakerBoehner
thenation	SpokesmanReview	NPR		taxreformer
thinkprogress	statesman	OpenSecretsDC		theamgreatness
Upworthy	Suntimes	PBS		theblaze
VICE	TeenVogue	pewresearch		TheDCPolitics
voxdotcom	TexasTribune	physorg_com		theMRC
yesmagazine	TheAtlantic	PressHerald		TheNatPulseRSS
	thedailynu	PsychScience		townhallcom
	TheEconomist	qz		TuckerCarlson
	TheJuanWilliams	Rasmussen_Poll		weeklystandard
	TheOnion	RealClearNews		WestJournalism
	TheRoot	Reuters		WhatfingerNews
	TIME	rollcall		worldnetdaily
	TimesCall	SatEvePost		
	TODAYshow	ScienceDaily		
	truthout	SFWeekly		
	UnivisionNews	snopes		
	urbaninstitute	TDOnline		
	usnews	TechCrunch		
	VanityFair	The_CUI		
	verge	thehill		
	washingtonpost	TheKoreaHerald		
	wigazette	TheNatlInterest		
	YahooNews	TheObserverNY		
		TheWeek		
		TheWorld		
		Timcast		
		USATODAY		
		Wikipedia		
		WIRED		
		WSJ		

Table A4: Media Bias Fact Check media outlets by Twitter handle and factual reporting

High		Mixed	Low
ABC	newrepublic	AlterNet	TheBabylonBee
ACLU	NYMag	AEI	BreitbartNews
AJEnglish	OpenSecretsDC	AmericanThinker	cnsnews
amnesty	PeacockPanache	BuzzFeedNews	FRCdc
AP	peoplefor	CNN	fpmag
ballotpedia	pewresearch	dailykos	JudicialWatch
business	politico	DailyMailUK	gatewaypundit
BoingBoing	politicususa	drudgefeed	TheNatPulseRSS
businessinsider	PolitiFact	FoxNews	TheOnion
CalWatchdog	TheWorld	GOP	
CarnegieEndow	propublica	IBTimes	
CBSNews	RANDCorporation	IBDinvestors	
amprog	RealClearNews	Jezebel	
publicintegrity	Reuters	theMRC	
CenterOnBudget	rollcall	MSNBC	
Suntimes	Salon	NRO	
chicagotribune	sfchronicle	NYDailyNews	
csmonitor	sciam	nypost	
CityJournal	SFGate	newsbusters	
CookPolitical	Slate	newsmax	
curaffairs	snopes	Newsweek	
thedailybeast	TechCrunch	OANN	
thedailynu	amconmag	PJMedia_com	
DailyProgress	amspectator	Project_Veritas	
DefenseOne	CJEducation	RawStory	
democracynow	TheEconomist	splinter_news	
EconomicPolicy	FDRLST	theblaze	
esquire	Independent	TheDCPolitics	
FAScientists	thenation	DailySignal	
FinancialTimes	NewYorker	realDailyWire	
TheFiscalTimes	TheObserverNY	Heritage	
ForeignAffairs	TPostMillennial	thehill	
GallupNews	TheRoot	WestJournalism	
grist	TexasTribune	theblaze	
TheIJR	verge	thinkprogress	
InsidePhilanthr	weeklystandard	townhallcom	
ivn	erumors	truthout	
jacobinmag	Upworthy	VICE	
lifehacker	USATODAY	FreeBeacon	
MarketWatch	VanityFair	WhiteHouse	
mashable	voxdotcom	worldnetdaily	
mmfa	WSJ	YahooNews	
Mediaite	dcexaminer		
MiamiHerald	monthly		
MotherJones	WashTimes		
ConstitutionCtr	Wikipedia		
nationaljournal	yesmagazine		

Table A5: Twitter: Effect of UI change, robustness checks

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Post	-0.178***	-0.176***	-0.263***	-0.072***	-0.101***	-0.203***	-0.157***
	(0.018)	(0.019)	(0.042)	(0.022)	(0.029)	(0.023)	(0.013)
N	284369	60384	284369	284369	284369	284369	515559
N-outlets (clusters)	137	137	137	137	137	137	332
R-sq	0.702	0.695	0.564	0.581	0.556	0.686	0.707
Outlet FE	Yes	Yes	No	Yes	Yes	Yes	Yes
Day FE	No	No	No	No	No	No	No
Excl. retweets	No	Yes	No	No	No	No	No
Likes (control)	Yes	Yes	Yes	No	No	Yes	Yes
Other controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
W. by bias agreement	No	No	No	No	Yes	Yes	No
Sample	Main	Main	Main	Main	Main	Main	Extnd.

Notes: OLS regressions with number of retweets (log) as the outcome. Sample includes tweets by selected media outlets in the 60-day window around the Twitter UI change. Controls include number of likes (log), whether the tweet contains a url link, a hashtag, an @ mention, whether the tweet is a reply, and tweet length. Standard errors clustered at the media outlet level. Significance levels indicated *p<0.10, ** p<0.05, ***p<0.01.

Table A6: Twitter: Effect of UI change by factual reporting, robustness checks

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Post	-0.137***									
	(0.031)									
Post x High-factual reporting	-0.064**	-0.062*	-0.065*	-0.078**	-0.118**	-0.118**	-0.079*	0.023	-0.014	0.056
	(0.032)	(0.032)	(0.037)	(0.037)	(0.048)	(0.053)	(0.040)	(0.044)	(0.041)	(0.038)
Post x Slant								0.204***		0.156***
								(0.059)		(0.040)
N	284369	284369	60384	284369	284369	284369	220335	220335	344097	344097
N-outlets (clusters)	137	137	137	137	137	137	71	71	136	136
R-sq	0.702	0.703	0.697	0.687	0.588	0.563	0.672	0.672	0.701	0.701
Outlet FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Excl. retweets	No	No	Yes	No	No	No	No	No	No	No
Likes (control)	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes
Other controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
W. by bias agreement	No	No	No	Yes	No	Yes	No	Yes	No	No
Factualness	MBFC	MBFC	MBFC	MBFC	MBFC	MBFC	AdFontes	AdFontes	Grinberg	Grinberg

Notes: OLS regressions with number of retweets (log) as the outcome. Sample includes tweets by selected media outlets in the 60-day window around the Twitter UI change. High-factualness is a dummy equal to 1 if the outlet has "high" factualness score. A description of the different factualness sources is in Appendix A.1. Controls include number of likes (log), whether the tweet contains a url link, a hashtag, an @ mention, whether the tweet is a reply, and tweet length. Standard errors clustered at the media outlet level. Significance levels indicated *p<0.10, ** p<0.05, ***p<0.01.

Table A7: Twitter: Effect of UI change by slant, robustness checks

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Post	-0.233***									-0.237***	-0.244***
	(0.019)									(0.024)	(0.019)
Post x Slant	0.127***	0.129***	0.122**	0.166***	0.192***	0.152***	0.136***	0.110**	0.216***		
	(0.042)	(0.042)	(0.051)	(0.060)	(0.064)	(0.045)	(0.033)	(0.043)	(0.064)		
Post x Outlet Slant (Lean left)										0.037	0.081***
										(0.030)	(0.028)
Post x Outlet Slant (Center/mixed)										0.089***	0.111***
										(0.034)	(0.024)
Post x Outlet Slant (Lean right)										0.073	0.125***
										(0.070)	(0.045)
Post x Outlet Slant (Right)										0.134***	0.146***
										(0.044)	(0.035)
N	284369	284369	60384	284369	284369	284369	508995	284369	168649	284369	515559
N-outlets (clusters)	137	137	137	137	137	137	329	137	54	137	332
R-sq	0.702	0.703	0.697	0.588	0.563	0.687	0.708	0.703	0.679	0.702	0.707
Outlet FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No
Excl. retweets	No	No	Yes	No	No	No	No	No	No	No	No
Likes (control)	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Other controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
W. by bias agreement	No	No	No	No	Yes	Yes	No	No	No	No	Yes
Sample	Main	Main	Main	Main	Main	Main	Extnd.	Main	Main	Main	Extnd.
Slant-measure	AllSides	AllSides	AllSides	AllSides	AllSides	AllSides	AllSides	MBFC	AdFontes	AllSides	AllSides

Notes: OLS regressions with number of retweets (log) as the outcome. Sample includes tweets by selected media outlets in the 60-day window around the Twitter UI change. Slant is defined as a continuous variable which varies from 0 (left) to 1 (right). A description of the different slant sources is in Appendix A.1. Controls include number of likes (log), whether the tweet contains a url link, a hashtag, an @ mention, whether the tweet is a reply, and tweet length. Standard errors clustered at the media outlet level. Significance levels indicated *p<0.10, ** p<0.05, ***p<0.01.

Table A8: Twitter: Effect of UI change by factual reporting and slant, robustness checks

	(1)	(2)	(3)	(4)	(5)	(6)
Post	-0.210***					
	(0.031)					
Post x High-factual reporting	-0.025	-0.021	-0.027	-0.053	-0.077*	-0.022
	(0.030)	(0.030)	(0.032)	(0.042)	(0.045)	(0.029)
Post x Slant	0.111***	0.115***	0.104**	0.159***	0.115**	0.138***
	(0.041)	(0.042)	(0.048)	(0.058)	(0.057)	(0.042)
N	284369	284369	60384	284369	284369	284369
N-outlets (clusters)	137	137	137	137	137	137
R-sq	0.702	0.703	0.697	0.563	0.588	0.687
Outlet FE	Yes	Yes	Yes	Yes	Yes	Yes
Day FE	No	Yes	Yes	Yes	Yes	Yes
Excl. retweets	No	No	Yes	No	No	No
Likes (control)	Yes	Yes	Yes	No	No	Yes
Other controls	Yes	Yes	Yes	Yes	Yes	Yes
W. by bias agreement	No	No	No	Yes	No	Yes

Notes: OLS regressions with number of retweets (log) as the outcome. Sample includes tweets by selected media outlets in the 60-day window around the Twitter UI change. Slant is defined as a continuous variable which varies from 0 (left) to 1 (right). Controls include number of likes (log), whether the tweet contains a url link, a hashtag, an @ mention, whether the tweet is a reply, and tweet length. Standard errors clustered at the media outlet level. Significance levels indicated *p<0.10, ** p<0.05, ***p<0.01.

Table A9: Twitter: Effect of UI change by slant (extreme), robustness checks

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Post	-0.175***										
	(0.021)										
Post x Extreme slant (Left or Right)	-0.008	-0.013	-0.012	-0.026	-0.058	-0.042	-0.038	-0.010	-0.035	-0.036	-0.009
	(0.030)	(0.030)	(0.035)	(0.043)	(0.049)	(0.035)	(0.024)	(0.028)	(0.037)	(0.031)	(0.028)
Post x High-factual reporting										-0.074**	
										(0.035)	
Post x Slant											0.129***
											(0.042)
N	284369	284369	60384	284369	284369	284369	515559	284369	284369	284369	284369
N-outlets (clusters)	137	137	137	137	137	137	332	137	137	137	137
R-sq	0.702	0.703	0.697	0.588	0.563	0.687	0.709	0.703	0.703	0.703	0.703
Outlet FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Excl. retweets	No	No	Yes	No	No	No	No	No	No	No	No
Likes (control)	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Other controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
W. by bias agreement	No	No	No	No	Yes	Yes	No	No	No	No	No
Sample	Main	Main	Main	Main	Main	Main	Extnd.	Main	Main	Main	Main
Slant-measure	AllSides	AllSides	AllSides	AllSides	AllSides	AllSides	AllSides	MBFC	AdFontes	AllSides	AllSides

Notes: OLS regressions with number of retweets (log) as the outcome. Sample includes tweets by selected media outlets in the 60-day window around the Twitter UI change. Extreme-slant is defined as a dummy equal to 1 if slant is "Left" or "Right". A description of the different slant sources is in Appendix A.1. Controls include number of likes (log), whether the tweet contains a url link, a hashtag, an @ mention, whether the tweet is a reply, and tweet length. Standard errors clustered at the media outlet level. Significance levels indicated *p<0.10, ** p<0.05, ***p<0.01.

Table A10: Facebook: Changes in sharing after policy intervention (election day), robustness checks

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Post	-0.125***	-0.188***	-0.138***	-0.163***	-0.205***	-0.150***	-0.140***
	(0.022)	(0.026)	(0.023)	(0.041)	(0.042)	(0.022)	(0.015)
N	196913	196913	196913	196913	196913	196913	333337
N-outlets (clusters)	137	137	137	137	137	137	299
R-sq	0.854	0.801	0.829	0.494	0.458	0.849	0.861
Outlet FE	Yes	Yes	No	Yes	Yes	Yes	Yes
Day FE	No	No	No	No	No	No	No
Eng. metrics	Yes	No	Yes	No	No	Yes	Yes
Likes (control)	Yes	Yes	Yes	No	No	Yes	Yes
Other controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
W. by bias agreement	No	No	No	No	Yes	Yes	No
Sample	Main	Main	Main	Main	Main	Main	Extnd.

Notes: OLS regressions with number of shares (log) as the outcome. Sample includes all posts by selected media outlets in the 60-day window around the FB policy change / election day. Controls include total engagement in logs (likes, wows, angry, etc), whether the post contains a url link, a hashtag, an @ mention, and post length. Standard errors clustered at the media outlet level. Significance levels indicated *p<0.10, ** p<0.05, ***p<0.01.

Table A11: Facebook: Changes in sharing after policy intervention (election day) by factual reporting, robustness checks

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Post	-0.191***									
	(0.034)									
Post x High-factual reporting	0.115***	0.114***	0.086*	0.119	0.108	0.111***	0.139***	0.102	0.131**	0.090
	(0.042)	(0.042)	(0.050)	(0.084)	(0.083)	(0.039)	(0.047)	(0.071)	(0.060)	(0.070)
Post x Slant								-0.069		-0.066
								(0.080)		(0.060)
N	196912	196912	196912	196912	196912	196912	157832	157832	258957	255331
N-outlets (clusters)	136	136	136	136	136	136	73	73	138	136
R-sq	0.854	0.855	0.802	0.499	0.463	0.850	0.848	0.848	0.853	0.853
Outlet FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Eng. metrics	Yes	Yes	No	No	No	Yes	Yes	Yes	Yes	Yes
Likes (control)	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes
Other controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
W. by bias agreement	No	No	No	No	Yes	Yes	No	No	No	No
Factualness	MBFC	MBFC	MBFC	MBFC	MBFC	MBFC	AdFontes	AdFontes	Grinberg	Grinberg

Notes: OLS regressions with number of shares (log) as the outcome. Sample includes all posts by selected media outlets in the 60-day window around the FB policy change / election day. High-factualness is a dummy equal to 1 if the outlet has "high" factualness score. A description of the different factualness sources is in Appendix A.1. Controls include total engagement in logs (likes, wows, angry, etc), whether the post contains a url link, a hashtag, an @ mention, and post length. Standard errors clustered at the media outlet level. Significance levels indicated *p<0.10, ** p<0.05, ***p<0.01.

Table A12: Facebook: Changes in sharing after policy intervention (election day) by slant, robustness checks

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Post	-0.068**									-0.114***	-0.116***
	(0.027)									(0.034)	(0.027)
Post x Slant	-0.130**	-0.131**	-0.005	-0.023	-0.033	-0.119**	-0.145***	-0.126**	-0.212**		
	(0.053)	(0.052)	(0.064)	(0.108)	(0.120)	(0.055)	(0.044)	(0.049)	(0.101)		
Post x Outlet Slant (Lean left)										0.039	0.011
										(0.041)	(0.032)
Post x Outlet Slant (Center/mixed)										0.036	0.011
										(0.078)	(0.050)
Post x Outlet Slant (Lean right)										-0.041	-0.023
										(0.050)	(0.036)
Post x Outlet Slant (Right)										-0.123**	-0.155***
										(0.056)	(0.047)
N	196913	196912	196912	196912	196912	196912	329355	196912	121457	196913	333337
N-outlets (clusters)	137	136	136	136	136	136	290	136	55	137	299
R-sq	0.854	0.855	0.802	0.499	0.463	0.850	0.862	0.855	0.844	0.854	0.862
Outlet FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No
Eng. metrics	Yes	Yes	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Likes (control)	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Other controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
W. by bias agreement	No	No	No	No	Yes	Yes	No	No	No	No	Yes
Sample	Main	Main	Main	Main	Main	Main	Extnd.	Main	Main	Main	Extnd.
Slant-measure	AllSides	AllSides	AllSides	AllSides	AllSides	AllSides	AllSides	MBFC	AdFontes	AllSides	AllSides

Notes: OLS regressions with number of shares (log) as the outcome. Sample includes all posts by selected media outlets in the 60-day window around the FB policy change / election day. Slant is defined as a continuous variable which varies from 0 (left) to 1 (right). A description of the different slant sources is in Appendix A.1. Controls include total engagement in logs (likes, wows, angry, etc), whether the post contains a url link, a hashtag, an @ mention, and post length. Standard errors clustered at the media outlet level. Significance levels indicated *p<0.10, ** p<0.05, ***p<0.01.

Table A13: Facebook: Changes in sharing after policy intervention (election day) by factual reporting and slant, robustness checks

	(1)	(2)	(3)	(4)	(5)	(6)
Post	-0.143***					
	(0.052)					
Post x High-factual reporting	0.093*	0.091*	0.100*	0.117	0.133	0.085**
	(0.049)	(0.049)	(0.059)	(0.084)	(0.089)	(0.036)
Post x Slant	-0.080	-0.082	0.049	0.025	0.049	-0.076
	(0.062)	(0.062)	(0.073)	(0.121)	(0.110)	(0.053)
N	196913	196912	196912	196912	196912	196912
N-outlets (clusters)	137	136	136	136	136	136
R-sq	0.854	0.855	0.802	0.463	0.499	0.850
Outlet FE	Yes	Yes	Yes	Yes	Yes	Yes
Day FE	No	Yes	Yes	Yes	Yes	Yes
Eng. metrics	Yes	Yes	No	No	No	Yes
Likes (control)	Yes	Yes	Yes	No	No	Yes
Other controls	Yes	Yes	Yes	Yes	Yes	Yes
W. by bias agreement	No	No	No	Yes	No	Yes

Notes: OLS regressions with number of shares (log) as the outcome. Sample includes all posts by selected media outlets in the 60-day window around the FB policy change / election day. Slant is defined as a continuous variable which varies from 0 (left) to 1 (right). Controls include total engagement in logs (likes, wows, angry, etc), whether the post contains a url link, a hashtag, an @ mention, and post length. Standard errors clustered at the media outlet level. Significance levels indicated *p<0.10, ** p<0.05, ***p<0.01.

Table A14: Facebook: Changes in sharing after policy intervention (election day) by slant (extreme), robustness checks

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Post	-0.089***										
	(0.028)										
Post x Extreme slant (Left or Right)	-0.082*	-0.080*	-0.052	-0.165**	-0.164**	-0.093**	-0.086***	-0.071*	-0.108	-0.032	-0.075*
	(0.042)	(0.042)	(0.050)	(0.077)	(0.075)	(0.037)	(0.033)	(0.043)	(0.072)	(0.042)	(0.040)
Post x High-factual reporting										0.099**	
										(0.044)	
Post x Slant											-0.126**
											(0.050)
N	196913	196912	196912	196912	196912	196912	333334	196912	196912	196912	196912
N-outlets (clusters)	137	136	136	136	136	136	296	136	136	136	136
R-sq	0.854	0.854	0.802	0.499	0.463	0.850	0.862	0.854	0.854	0.855	0.855
Outlet FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Eng. metrics	Yes	Yes	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Likes (control)	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Other controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
W. by bias agreement	No	No	No	No	Yes	Yes	No	No	No	No	No
Sample	Main	Main	Main	Main	Main	Main	Extnd.	Main	Main	Main	Main
Slant-measure	AllSides	AllSides	AllSides	AllSides	AllSides	AllSides	AllSides	MBFC	AdFontes	AllSides	AllSides

Notes: OLS regressions with number of shares (log) as the outcome. Sample includes all posts by selected media outlets in the 60-day window around the FB policy change / election day. Extreme-slant is defined as a dummy equal to 1 if slant is "Left" or "Right". A description of the different slant sources is in Appendix A.1. Controls include total engagement in logs (likes, wows, angry, etc), whether the post contains a url link, a hashtag, an @ mention, and post length. Standard errors clustered at the media outlet level. Significance levels indicated *p<0.10, ** p<0.05, ***p<0.01.

Table A15: Twitter: Effect of UI change with baseline controls

	(1)	(2)	(3)	(4)
Post x Baseline shares	-0.005	-0.006	-0.007	-0.003
	(0.014)	(0.013)	(0.013)	(0.014)
Post x High-factual reporting	-0.069*		-0.020	-0.015
	(0.035)		(0.036)	(0.038)
Post x Slant		0.137***	0.123**	0.110**
		(0.044)	(0.048)	(0.055)
N	183433	183433	183433	39808
N-outlets (clusters)	111	111	111	111
R-sq	0.702	0.703	0.703	0.695
Outlet FE	Yes	Yes	Yes	Yes
Day FE	Yes	Yes	Yes	Yes
Excl. retweets	No	No	No	Yes
Likes (control)	Yes	Yes	Yes	Yes
Other controls	Yes	Yes	Yes	Yes

Notes: OLS regressions with number of retweets (log) as the outcome. Sample includes all posts by selected media outlets in the 60-day window around the Twitter policy change / election day. Slant is defined as a continuous variable which varies from 0 (left) to 1 (right). High-factualness is a dummy equal to 1 if the outlet has "high" factualness score. Controls include total engagement in logs (likes), whether the post contains a url link, a hashtag, an @ mention, and tweet length. Baseline shares are the mean number of log retweets during the first week of October 2020. Standard errors clustered at the media outlet level. Significance levels indicated *p<0.10, ** p<0.05, ***p<0.01.

Table A16: Facebook: Changes in sharing after policy intervention (election day) with baseline controls

	(1)	(2)	(3)	(4)	(5)
Post x Baseline shares	-0.042**	-0.051***	-0.013	-0.020	-0.012
	(0.016)	(0.017)	(0.034)	(0.031)	(0.032)
Post x High-factual reporting	0.054	0.117***	0.084*		0.069
	(0.046)	(0.043)	(0.042)		(0.046)
Post x Slant				-0.078	-0.045
				(0.057)	(0.061)
N	196910	157832	112902	112902	112902
N-outlets (clusters)	135	73	65	65	65
R-sq	0.855	0.849	0.811	0.811	0.811
Outlet FE	Yes	Yes	Yes	Yes	Yes
Day FE	Yes	Yes	Yes	Yes	Yes
Eng. metrics	Yes	Yes	Yes	Yes	Yes
Likes (control)	Yes	Yes	Yes	Yes	Yes
Other controls	Yes	Yes	Yes	Yes	Yes
Factualness	MBFC	AdFontes	MBFC	MBFC	MBFC

Notes: OLS regressions with number of shares (log) as the outcome. Sample includes all posts by selected media outlets in the 60-day window around the FB policy change / election day. Slant is defined as a continuous variable which varies from 0 (left) to 1 (right). High-factualness is a dummy equal to 1 if the outlet has "high" factualness score. Controls include total engagement in logs (likes, wows, angry, etc), whether the post contains a url link, a hashtag, an @ mention, and post length. Baseline shares are the mean number of log shares during the first week of October 2020. Standard errors clustered at the media outlet level. Significance levels indicated *p<0.10, ** p<0.05, ***p<0.01.