SOCIAL MEDIA

Reshares on social media amplify political news but do not detectably affect beliefs or opinions

Andrew M. Guess[1]*, Neil Malhotra[2], Jennifer Pan[3], Pablo Barberá[4], Hunt Allcott[5], Taylor Brown[4], Adriana Crespo-Tenorio[4], Drew Dimmery[4,6], Deen Freelon[7], Matthew Gentzkow[8], Sandra González-Bailón[9], Edward Kennedy[10], Young Mie Kim[11], David Lazer[12], Devra Moehler[4], Brendan Nyhan[13], Carlos Velasco Rivera[4], Jaime Settle[14], Daniel Robert Thomas[4], Emily Thorson[15], Rebekah Tromble[16], Arjun Wilkins[4], Magdalena Wojcieszak[17,18], Beixian Xiong[4], Chad Kiewiet de Jonge[4], Annie Franco[4], Winter Mason[4], Natalie Jomini Stroud[19], Joshua A. Tucker[20]

We studied the effects of exposure to reshared content on Facebook during the 2020 US election by assigning a random set of consenting, US-based users to feeds that did not contain any reshares over a 3-month period. We find that removing reshared content substantially decreases the amount of political news, including content from untrustworthy sources, to which users are exposed; decreases overall clicks and reactions; and reduces partisan news clicks. Further, we observe that removing reshared content produces clear decreases in news knowledge within the sample, although there is some uncertainty about how this would generalize to all users. Contrary to expectations, the treatment does not significantly affect political polarization or any measure of individual-level political attitudes.

It has been more than 10 years since social media platforms began to adopt reposting as a core feature. More than one-fourth of posts that Facebook users see in their feeds have been "reshared" [see supplementary materials (SM) section S1.8]: Rather than originating from a friend, a followed page, or a group they belong to, these posts appear in their feeds when they are reposted by someone to whom they are directly connected. For Facebook and other platforms that enable reshare functionality, content rapidly attains popularity, or "goes viral," because it is reshared (1–5), and removing reshared content has been suggested as a policy intervention to minimize harmful effects of virality (6–8). However, the effects of such an intervention are unknown because most reshared content does not go viral. Given platforms' importance as facilitators of public discourse, this key element of the social media experience could affect political attitudes and behaviors. Accordingly, we investigate the impact of removing reshared content from the feeds of consenting Facebook users on political polarization and political knowledge with a randomized controlled experiment conducted within the context of the 2020 US presidential election campaign.

Theory and research questions

Prior research suggests that reshared content may increase political polarization and political knowledge. For example, experimental research has shown that deactivating Facebook causes a decrease in both polarization and knowledge (9), suggesting that these outcomes may be related. Resharing is a key feature of social platforms that could plausibly drive these effects. In particular, reshared posts with strong social endorsements seem more credible than non-reshared posts and may thus be more noticeable to users (10, 11). Users tend to reshare others' content when it is emotionally activating (12), including politically tinged content that could feed polarized attitudes (13–16).

Withholding reshared content from users' feeds may thus reduce affective and issue polarization by decreasing their exposure to emotionally or ideologically inflammatory content (H1). The hypothesized mechanism is that reshares are likely to promote encounters with partisan cross-cutting sources, increasing the potential for engagement with polarizing content. Some experimental research suggests that exposure to such content could increase issue polarization (17), though evidence is still mixed (14). By contrast, reshared content may have a net benefit for political knowledge. Though reshares can increase exposure to misinformation and lower-quality content (e.g., clickbait, sharebait) (4, 5), they may also incidentally expose users to trustworthy information they do not seek out, facilitating by-product learning about the election and events in the news (18–22). Given that misinformation is a small fraction of content on social platforms (23, 24), we expect that removing reshared content would, on net, reduce accurate knowledge about the election campaign (H2). We additionally test a series of secondary hypotheses about the effect of withholding reshared content on a range of outcomes including factual discernment and belief in the integrity of the 2020 election.

Design and results

As part of the US 2020 Facebook and Instagram Election Study (FIES), participants were recruited through survey invitations placed on the top of their Facebook feeds beginning in August 2020, which were seen by ~14.6 million users. Adult users residing in the United States who provided informed consent (N = 193,880, 1.3% of those who saw the invitation) were invited to complete five surveys and share their Facebook activity (see SM section S1.1 for study timeline and CONSORT flow diagram). Participants were given the option to withdraw from the study and/or withdraw their data from the study up until the data were disconnected from any identifiers in February 2023 (see SM section S8.2 for additional details on recruitment and consent, and sections S1.2 and S12 for ethical considerations). Participants who completed baseline surveys (N = 75,189, 0.5% of those who saw the invitation) were randomized into eight experimental conditions associated with the larger FIES study. These participants were separately invited to share data on their desktop and mobile web visits (N = 7730 consented and shared data, or 0.05% of those who saw the initial study invitation).

In this article, we analyze data from two of the experimental conditions: (i) a control condition in which no changes were made to their Facebook feeds, and (ii) a treatment condition in which no reshared content (from friends, Groups, or Pages) was shown in the feed (for simplicity, we refer to this as the No Reshares group). Both groups were compensated equally for participation. The experimental intervention was active from 24 September to 23 December 2020. Our samples (N = 23,402 for survey data and N = 3781 for passive tracking data), comprising participants who completed the first two surveys and at least one of the subsequent three waves, allow for adequate statistical power to detect small effects (e.g., for affective polarization, we are powered to detect population average treatment effects with Cohen's d = 0.032 or larger and sample average treatment effects d = 0.023 or larger). Although

[1]Department of Politics and School of Public and International Affairs, Princeton University, Princeton, NJ, USA. [2]Graduate School of Business, Stanford University, Stanford, CA, USA. [3]Department of Communication, Stanford University, Stanford, CA, USA. [4]Meta, Menlo Park, CA, USA, CA, USA. [5]Stanford Doerr School of Sustainability, Stanford University, Stanford, CA, USA. [6]Research Network Data Science, University of Vienna, Vienna, Austria. [7]UNC Hussman School of Journalism and Media, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [8]Department of Economics, Stanford University, Stanford, CA, USA. [9]Annenberg School for Communication, University of Pennsylvania, Philadelphia, PA, USA. [10]Department of Statistics and Data Science, Carnegie Mellon University, Pittsburgh, PA, USA. [11]School of Journalism and Mass Communication, University of Wisconsin-Madison, Madison, WI, USA. [12]Network Science Institute, Northeastern University, Boston, MA, USA. [13]Department of Government, Dartmouth College, Hanover, NH, USA. [14]Department of Government, William & Mary, Williamsburg, VA, USA. [15]Department of Political Science, Syracuse University, Syracuse, NY, USA. [16]School of Media and Public Affairs and Institute for Data, Democracy, and Politics, The George Washington University, Washington, DC, USA. [17]Department of Communication, University of California, Davis, CA, USA. [18]Amsterdam School of Communication Research, University of Amsterdam, Amsterdam, Netherlands. [19]Moody College of Communication and Center for Media Engagement, University of Texas at Austin, Austin, TX, USA. [20]Wilf Family Department of Politics and Center for Social Media and Politics, New York University, New York, NY, USA.
*Corresponding author. Email: aguess@princeton.edu

technical constraints meant that respondents in the No Reshares group saw a small amount of reshared content (see S1.8 for more details), respondents in the No Reshares group saw substantially less reshared content than those in the control group: 28% of views in the control group were of reshared content, compared with 5.8% in the No Reshares treatment group ($p < 0.01$).

Among participants in the experimental sample, 19.5% did not complete any of our posttreatment survey waves. However, this attrition is not significantly different between treatment and control groups ($p = 0.75$). We also did not observe any differential attrition by survey wave across the study period (see S1.9). For additional information on these and other aspects of the study design, see SM section S1, materials and methods. The study design, measures, and analysis were preregistered at the Open Science Framework (OSF) prior to treatment assignment. As discussed in the preanalysis plan, our main estimand of interest is the population average treatment effect (PATE), which is weighted by users' predicted ideology, friend

count, number of political pages followed, and number of days active, among other variables (see SM section S8.5). We also report the unweighted sample average treatment effect (SATE) among consenting participants for transparency. The unweighted sample is more active on the platform, more likely to be between the ages of 30 and 44, more white, more female, more liberal, higher income, and more likely to have a college degree than the weighted sample (see SM section S3.3 for detailed comparisons). Our PATE estimates are designed to facilitate inferences to the population assuming negligible treatment effect heterogeneity by other characteristics. In general, and consistent with (25), we find limited effect heterogeneity (see S2.3). In case our weighting scheme does not fully account for the greater activity levels in the sample, our estimates correspond to arguably the most relevant subset of users—those who engage the most and generate a disproportionate share of content on the platform (see S2.1).

Removing reshared content substantially altered how participants used Facebook and the

content they saw (Fig. 1). Participants in this study spent more time on Facebook than the average US monthly average user (see SM section S3.3 for details), but the control group spent 73% more time each day on average compared to US monthly active users, whereas time spent reduced to 64% more for those in the No Reshares group ($p < 0.005$, Fig. 1A; all reported p-values are from covariate-adjusted ordinary least squares regression models with HC2 robust standard errors). We do not observe significant substitution to other social media platforms (see SM section S2.1). Users clicked on 7.5% of content they saw in the No Reshares group as opposed to 8.3% in the control group ($p = 0.04$, Fig. 1B). The proportion of posts seen by respondents to which they added any reaction (love, sadness, anger, surprise, or caring) also decreased ($p < 0.01$), but there were no significant changes in the rate of likes, comments, or users' own reshares (see Table S25 in SM section S2.1 for more details). This means that removing reshared content does not change users' own reshare behavior,
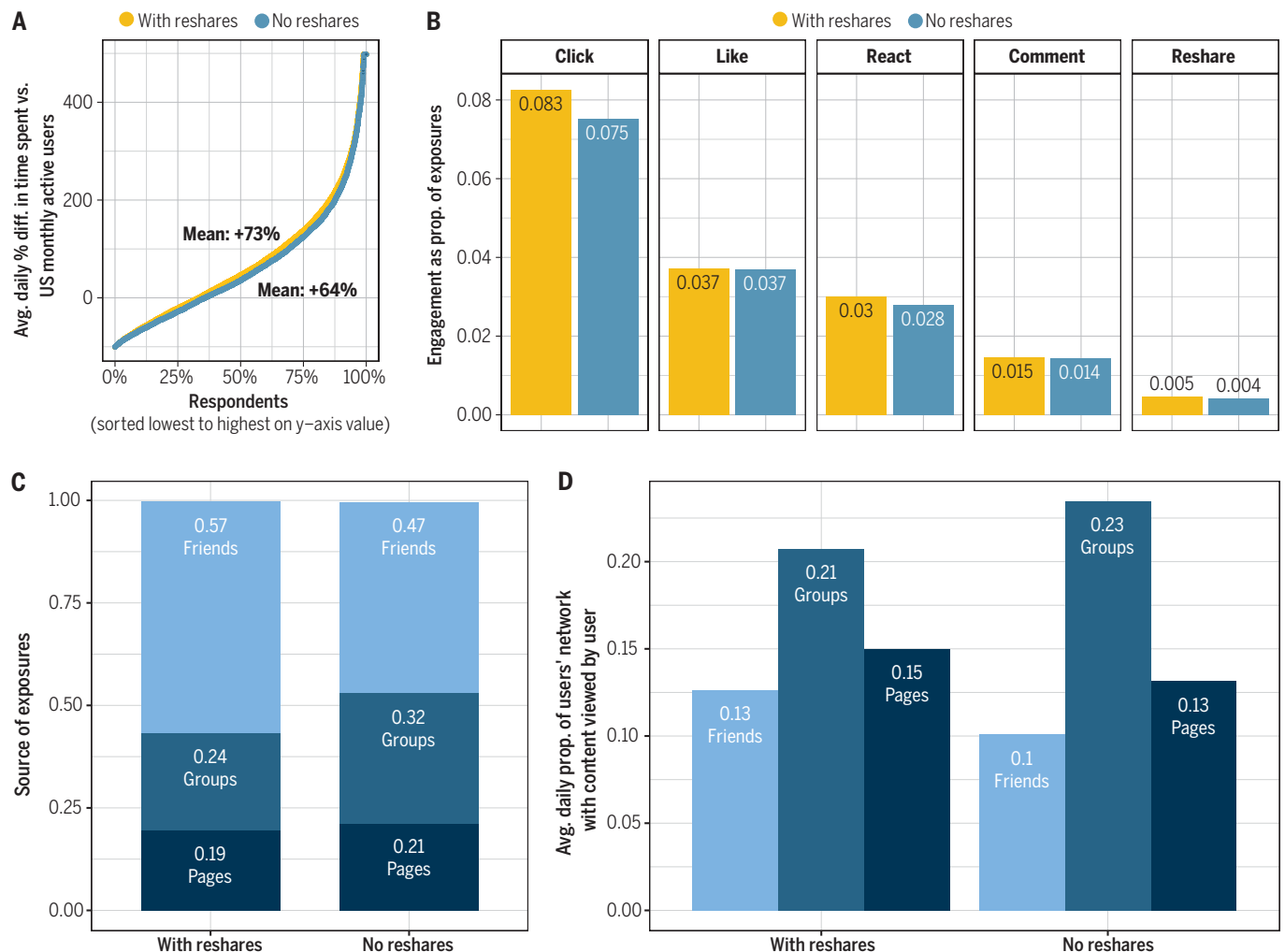
Fig. 1. Comparison of user experience and behavior for the No Reshares treatment and control conditions. (**A** to **D**) Values are unweighted sample statistics ($N = 23{,}402$). All differences are significant at the $p < 0.005$ level except for panel B Click rate ($p < 0.05$), Like rate ($p > 0.05$), Comment rate ($p > 0.05$), and Reshare rate ($p > 0.05$); confidence intervals are thus not shown.

which aligns with our understanding that most reshared content does not go viral: Resharing content typically does not create subsequent chains and cascades of resharing activity that result in virality. The No Reshares treatment decreased the relative proportion of content seen by participants that is posted by their friends by an average of 10 percentage points while increasing the relative share of content from Groups by 8 percentage points and from Pages by 2 percentage points ($p < 0.01$ for all comparisons, Fig. 1C). The No Reshares group decreased the proportion of users' networks of friends and Pages from which they saw content, but the share from users' network of Groups increased ($p < 0.01$ for all comparisons, Fig. 1D).

Users in the No Reshares treatment group saw a different mix of content in their feeds, with the largest changes occurring in political news, untrustworthy sources, and content classified as uncivil, as well as political content more generally (see S7 for more details on classification methods). As shown in Fig. 2, feeds without reshares contained less content pertaining to politics on average compared to the control group (10.8 versus 13.5%, $p < 0.005$), a difference that was especially pronounced for political news content, which decreased by more than half, on average (2.5 versus 6.2%, $p < 0.005$). Suppressing reshares cut the share of content from untrustworthy sources by nearly a third relative to the control group (1.8 versus 2.6%, $p < 0.005$), whereas it increased exposure to posts classified as uncivil by more than 6% (3.4 versus 3.2%, $p < 0.005$). Overall, however, the levels of exposure to both content from untrustworthy sources and uncivil posts were relatively low at baseline. We also classified content from other users, Pages, and Groups in participants' feeds as ideologically "like-minded" or "cross-cutting" (see SM section S1, materials and methods). The No Reshares condition decreased the proportion of both like-minded (51.1 versus 53.7%, $p < 0.005$) and cross-cutting (19.7 versus 20.7%, $p < 0.005$) content, while increasing that of ideologically moderate content by more than 15% (26.2 versus 22.6%, $p < 0.005$). Our analysis of treatment effect heterogeneity reveals that the decreases in political and political news content were greatest among those with the highest pre-treatment levels of Facebook activity, but effects mostly did not vary by other prespecified subgroups (see S2.3 for all subgroup analyses). Finally, exposure to speech with slur words, which was extremely rare at baseline, was not significantly affected by our treatment.

Turning to tests of our primary hypotheses (Table 1, top section), we find that respondents without reshared content in their feeds did not express significantly lower levels of affective or issue polarization than those in the control group [population average treatment effect: false discovery rate (FDR) adjusted $p > 0.8$ in both cases], meaning that we do not find support for H1. There was also no statistically distinguishable change in election knowledge, i.e., users were less likely to correctly remember recent events ($p > 0.8$). However, focusing on the sample average treatment effect suggests that removing reshares significantly reduced news knowledge, i.e., users were less likely to correctly remember recent events ($p < 0.01$). When estimating the population average treatment effect to generalize beyond the consenting participant sample to the population of Facebook users, the estimate for news knowledge falls short of our preregistered threshold for statistical significance when applying our correction for multiple testing ($p = 0.16$). According to our primary estimand and prespecified threshold, H2 is also not supported. Yet, the effect sizes are fairly similar between the sample average treatment effect and the population average treatment effect. Hence, removing reshared content produces clear decreases in news knowledge within the sample, although there is some uncertainty about how this would generalize to all users (see S3.3 for comparisons between the weighted and unweighted samples).

In terms of secondary hypotheses (Table 1, bottom section), we observe no differences between the treatment and control conditions that survive our preregistered correction for multiple comparisons (26) across nearly all outcomes (see SM section S1.11). The treatment does not have statistically distinguishable effects on perceived accuracy of various factual claims, trust in media (either traditional or social), confidence in political institutions, perceptions of political polarization, political efficacy, belief in the legitimacy of the election, or support for political violence. We also do not detect differences in the consumption of political news off platform as measured with behavioral web visit data; although this may be surprising given the shifts in content that we observe, it is consistent with the low overall proportion of political news in people's feeds. The one notable exception is that users in the No Reshares condition clicked less frequently on political news content from likely partisan sources (population average treatment effect: −0.109 SD, $p < 0.01$).

For both the primary and secondary hypotheses, there are two results for the population average treatment effect that individually fall
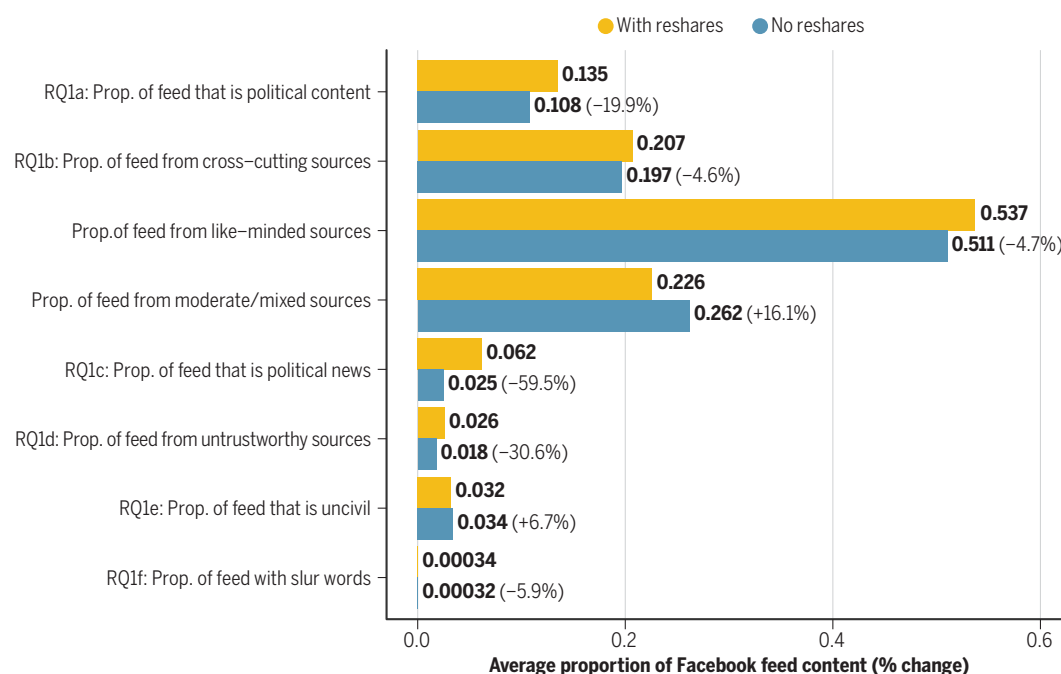

Fig. 2. Estimated changes in prevalence of feed content on Facebook. Values are average unweighted proportions within each group, with percent changes relative to the control group in parentheses (N = 23,402). All differences are significant at the $p < 0.005$ level, except RQ1f ($p > 0.05$); confidence intervals are thus not shown. Fully specified regression models with survey weights are reported in SM section S2.2.

Legend: ● With reshares ● No reshares

RQ1a: Prop. of feed that is political content — 0.135 / 0.108 (−19.9%)
RQ1b: Prop. of feed from cross–cutting sources — 0.207 / 0.197 (−4.6%)
Prop. of feed from like–minded sources — 0.537 / 0.511 (−4.7%)
Prop. of feed from moderate/mixed sources — 0.226 / 0.262 (+16.1%)
RQ1c: Prop. of feed that is political news — 0.062 / 0.025 (−59.5%)
RQ1d: Prop. of feed from untrustworthy sources — 0.026 / 0.018 (−30.6%)
RQ1e: Prop. of feed that is uncivil — 0.032 / 0.034 (+6.7%)
RQ1f: Prop. of feed with slur words — 0.00034 / 0.00032 (−5.9%)

Average proportion of Facebook feed content (% change)

Table 1. Population average treatment effects (PATE) and sample average treatment effects (SATE) of the No Reshares group (relative to the control group) for primary hypotheses (top) and secondary hypotheses (bottom). Estimates are presented in standard deviations (SD), and *p*-values are presented with and without sharpened FDR adjustment (*26*). Outcomes marked with * were computed using passive tracking data subsample (*N* = 3,781; see SM section S9). Outcomes marked with † were computed using Facebook platform data (*N* = 23,402; see SM section S6). All other outcomes were computed using survey data (*N* = 23,402; see SM section S1.5).

Hypothesis	Estimand	ATE	95% CI	SE	*p*	*p*adj
H1a: Affective polarization	PATE	−0.006	[−0.034, 0.022]	0.014	0.673	1.000
	SATE	0.012	[−0.005, 0.030]	0.009	0.158	0.188
H1b: Issue polarization	PATE	0.012	[−0.018, 0.042]	0.015	0.418	0.895
	SATE	0.007	[−0.010, 0.025]	0.009	0.414	0.268
H2a: Election knowledge	PATE	−0.017	[−0.064, 0.030]	0.024	0.472	0.895
	SATE	−0.021	[−0.048, 0.006]	0.014	0.124	0.188
H2b: News knowledge	PATE	−0.053	[−0.101, −0.004]	0.025	0.034	0.157
	SATE	−0.069	[−0.095, −0.043]	0.013	<0.01	<0.01
SH1a: Factual discernment	PATE	−0.045	[−0.082, −0.008]	0.019	0.016	0.151
	SATE	−0.027	[−0.048, −0.005]	0.011	0.015	0.083
SH1b: Off-platform political news visits*	PATE	0.061	[−0.042, 0.164]	0.052	0.243	0.751
	SATE	0.044	[−0.022, 0.111]	0.034	0.193	0.348
SH2a: Trust in media (excluding social)	PATE	−0.015	[−0.042, 0.012]	0.014	0.264	0.751
	SATE	−0.017	[−0.033, −0.002]	0.008	0.031	0.122
SH2b: Trust in info from social media	PATE	0.017	[−0.020, 0.053]	0.019	0.371	0.782
	SATE	−0.013	[−0.035, 0.009]	0.011	0.245	0.416
SH2c: Confidence in institutions	PATE	0.021	[−0.024, 0.067]	0.023	0.354	0.782
	SATE	0.010	[−0.015, 0.035]	0.013	0.445	0.506
SH3a: Perceived polarization	PATE	−0.028	[−0.066, 0.011]	0.020	0.157	0.579
	SATE	−0.008	[−0.029, 0.014]	0.011	0.480	0.506
SH3b: Partisan news clicks†	PATE	−0.109	[−0.159, −0.059]	0.025	<0.01	<0.01
	SATE	−0.105	[−0.128, −0.082]	0.012	<0.01	<0.01
SH3c: Partisan news visits*	PATE	−0.049	[−0.152, 0.054]	0.052	0.348	0.782
	SATE	0.004	[−0.066, 0.074]	0.036	0.900	0.713
SH4: Epistemic political efficacy	PATE	0.032	[−0.006, 0.070]	0.019	0.098	0.521
	SATE	0.017	[−0.006, 0.039]	0.011	0.146	0.311
SH5a: Party-line presidential voting	PATE	0.002	[−0.033, 0.037]	0.018	0.920	1.000
	SATE	−0.004	[−0.022, 0.014]	0.009	0.655	0.575
SH5b: Party-line down-ballot voting	PATE	0.012	[−0.036, 0.060]	0.025	0.625	0.966
	SATE	0.012	[−0.016, 0.040]	0.014	0.391	0.506
SH6: Belief in legitimacy of the election	PATE	−0.026	[−0.061, 0.009]	0.018	0.143	0.579
	SATE	−0.019	[−0.038, 0.001]	0.010	0.058	0.176
SH7: Support for political violence	PATE	−0.008	[−0.051, 0.035]	0.022	0.718	0.976
	SATE	−0.003	[−0.029, 0.024]	0.014	0.840	0.713

below the threshold for statistical significance but cross above the 0.05 preregistered threshold when multiple comparisons adjustments are implemented. In addition to having lower news knowledge on average, participants in the No Reshares group were less likely to be able to discern false claims. These two outcomes are related and substantively important, which prompted us to conduct non-preregistered, exploratory analyses to probe potential mechanisms that may explain these results (see S3). The evidence that we explore suggests that several mechanisms are unlikely to play a role. Specifically, we do not find that the No Reshares treatment increased exposure to Groups in which users frequently share untrustworthy information or decreased exposure to Facebook products aimed at increasing general knowledge (e.g., Voter Hub views); and, as noted previously, it did not increase exposure to like-minded sources. We do see a 62% decrease in exposure to mainstream news sources in the No Reshares group, which may have contributed to these results. Given the exploratory nature of these analyses, caution is warranted in interpreting these additional results and possible explanations.

Discussion

Content that goes viral on Facebook does so primarily through the mechanism of reshares. But the kinds of cascades typically associated with viral memes and emotionally engaging stories are rare (*27*). More often, reshares simply promote greater exposure to content from Facebook friends (relative to Groups) in a user's network and increase overall engagement with posts (in the form of clicks and reactions). By removing reshared posts from users' feeds, our experiment reveals the average impact of reshares on the mix of content that they see. The two content types we measured that changed the most—political news and posts from untrustworthy sources—suggest that the reshares feature could be a double-edged sword: It facilitates encounters with both reliable news about politics and current events but, to a somewhat lesser degree, also content from untrustworthy sources that may exaggerate or fabricate information.

However, despite these changes, we were not able to reliably detect shifts in users' political attitudes or behaviors, with the exception of a decrease in news knowledge within the sample. Owing to the large size of our experiment, we can rule out even moderately sized effects. This disconnect is illustrated by the one outcome among our secondary hypotheses on which we estimate a statistically distinguishable effect of the No

Reshares treatment: a decrease in partisan news clicks. Our finding indicates that without reshares in their feeds, users were less likely to click on outbound links from news sources with highly ideological audiences. Yet this clear example of reshares-driven on-platform behavior does not manifest in our survey measures of issue or affective polarization. We conclude that though reshares may have been a powerful mechanism for directing users' attention and behavior on Facebook during the 2020 election campaign, they had limited impact on politically relevant attitudes and offline behaviors.

Although our study provides rigorous evidence of the effects of reshared content on political outcomes, our conclusions are limited to the period in which we conducted our experiment—relatively late in terms of user adoption and in the midst of a politically divisive period in American history. Finally, our design cannot speak to "general equilibrium" effects, because doing so would imply making inferences about societal impact—for example, how the demand for certain kinds of content, and consequently, the incentives of content producers might change—were our randomized intervention scaled up to the population of all users. Nonetheless, our findings lay the groundwork for future research to investigate these nuances.

REFERENCES AND NOTES

1. E. Bakshy, I. Rosenn, C. Marlow, L. Adamic, in *Proceedings of the 21st International Conference on World Wide Web* (2012), pp. 519–528.
2. J. Cheng, L. Adamic, P. A. Dow, J. M. Kleinberg, J. Leskovec, in *Proceedings of the 23rd International Conference on World Wide Web* (2014), pp. 925–936.
3. J. Ugander, B. Karrer, L. Backstrom, C. Marlow, The Anatomy of the Facebook Social Graph. arXiv:1111.4503 [cs.SI] (2011).
4. A. Friggeri, L. Adamic, D. Eckles, J. Cheng, in *Proceedings of the International AAAI Conference on Web and Social Media* (2014), vol. 8, pp. 101–110.
5. S. Vosoughi, D. Roy, S. Aral, *Science* **359**, 1146–1151 (2018).
6. A. Kantrowitz, "The Man Who Built The Retweet: 'We Handed A Loaded Weapon To 4-Year-Olds,'" *BuzzFeed News* (2019); https://www.buzzfeednews.com/article/alexkantrowitz/how-the-retweet-ruined-the-internet.
7. C. Newton, *Platformer* (2020); https://www.platformer.news/p/twitter-adds-friction.
8. J. Haidt, "Why the Past 10 Years of American Life Have Been Uniquely Stupid," *Atlantic* (2022); https://www.theatlantic.com/magazine/archive/2022/05/social-media-democracy-trust-babel/629369/.
9. H. Allcott, L. Braghieri, S. Eichmeyer, M. Gentzkow, *Am. Econ. Rev.* **110**, 629–676 (2020).
10. S. Messing, S. J. Westwood, *Communic. Res.* **41**, 1042–1063 4 (2014).
11. K. Munger, *Polit. Commun.* **37**, 376–397 (2020).
12. J. Berger, K. L. Milkman, *J. Mark. Res.* **49**, 192–205 (2012).
13. W. J. Brady, J. A. Wills, J. T. Jost, J. A. Tucker, J. J. Van Bavel, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 7313–7318 (2017).
14. R. Levy, *Am. Econ. Rev.* **111**, 831–870 (2021). doi:10.1257/aer.20191777
15. G. Pennycook *et al.*, *Nature* **592**, 590–595 (2021).
16. M. Wojcieszak, A. Casas, X. Yu, J. Nagler, J. A. Tucker, *Sci. Adv.* **8**, eabn9418 (2022).
17. C. A. Bail *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 9216–9221 (2018).
18. M. Prior, *Post-Broadcast Democracy: How Media Choice Increases Inequality in Political Involvement and Polarizes Elections* (Cambridge University Press, 2007).
19. L. Bode, *Mass Commun. Soc.* **19**, 24–48 (2016).
20. R. Fletcher, R. K. Nielsen, *New Media Soc.* **20**, 2450–2468 (2018).
21. J. T. Feezell, B. Ortiz, *Inf. Commun. Soc.* **24**, 1283–1302 (2021).
22. A. Acerbi, S. Altay, H. Mercier, Research note: Fighting misinformation or fighting for information? *Harvard Kennedy School (HKS) Misinformation Review* (2022); https://doi.org/10.37016/mr-2020-87.
23. J. Allen, B. Howland, M. Mobius, D. Rothschild, D. J. Watts, *Sci. Adv.* **6**, eaay3539 (2020).
24. A. Guess, K. Aslett, J. Tucker, R. Bonneau, J. Nagler, *J. Quant. Descr. Digit. Media* 10.51685/jqd.2021.006 (2021).
25. A. Coppock, *Political Sci. Res. Methods* **7**, 613–628 (2019).
26. Y. Benjamini, A. M. Krieger, D. Yekutieli, *Biometrika* **93**, 491–507 (2006).
27. P. A. Dow, L. Adamic, A. Friggeri, in *Proceedings of the International AAAI Conference on Web and Social Media* (2013), vol. 7, pp. 145–154.

SUPPLEMENTARY MATERIALS