Journal of Experimental Psychology: General

Do People Prefer to Share Political Information That Boosts Their Ingroup or Derogates the Outgroup?

Jakob Kasper and Thomas Gilovich

Online First Publication, January 27, 2025. https://dx.doi.org/10.1037/xge0001683

CITATION

Kasper, J., & Gilovich, T. (2025). Do people prefer to share political information that boosts their ingroup or derogates the outgroup? *Journal of Experimental Psychology: General*. Advance online publication. https://dx.doi.org/10.1037/xge0001683

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

© 2025 American Psychological Association
ISSN: 0096-3445

https://doi.org/10.1037/xge0001683

Do People Prefer to Share Political Information That Boosts Their Ingroup or Derogates the Outgroup?

Jakob Kasper[1, 2, 3] and Thomas Gilovich[2]
[1] Department of Psychology, Heidelberg University
[2] Department of Psychology, Cornell University
[3] Amsterdam School of Communication Research, University of Amsterdam

Recent analyses of social media activity indicate that outgroup animosity drives user engagement more than ingroup favoritism, with content that derogates the outgroup tending to generate more viral responses online. However, it is unclear whether those findings are due to most people's underlying preferences or structural features of the social media landscape. To address this uncertainty, we conducted three experimental studies ($N_{\text{overall}} = 609$) to examine how intended impact (ingroup favoritism/outgroup derogation) influences intentions to share both true and false news posts among U.S. partisans who regularly use social media. Participants consistently preferred to share posts that favor their own party over those that denigrate the opposition—a preference that was largely maintained despite a manipulation of ingroup threat or a manipulated desire to share viral content in Studies 2 and 3. We discuss the influence of polarized politicians and their followers, malign actors, and social media algorithms as potential drivers of earlier results that highlight the virality of derogatory content.

Public Significance Statement
Social media platforms are important sources of political information that influence contemporary political debates. It is therefore a significant concern that recent analyses have documented the viral spread of outgroup hostility on these platforms. Here, we report three studies showing that U.S. political partisans generally prefer to share content that extols their political ingroup rather than denigrates the outgroup. This preference persisted even when participants were exposed to threats to their own group, and they did not show a preference for derogatory content when given the aim of sharing viral content. Although our findings do not call into question the prevalence and virality of outgroup derogation online, they support a different conception of the average social media user—one that can have beneficial societal consequences by highlighting ways to curb the spread of inflammatory content (e.g., by reducing the influence of hyperpartisan minorities and polarizing algorithms).

Keywords: social media, polarization, social identity, ingroup favoritism, outgroup derogation

Much has been written about the recent increase in political polarization, especially in the United States (Finkel et al., 2020; Iyengar et al., 2019; Van Bavel, Rathje, et al., 2021). It is a topic that warrants attention because extreme polarization can spark violence and erode democracy and because it makes it harder to achieve the consensus necessary to address pressing societal problems (Druckman et al., 2013; Iyengar et al., 2019; Lorenz-Spreen et al., 2023; Munn, 2021; Whitt et al., 2021).

Analyses of contemporary polarization often distinguish between (a) *ideological polarization*, or divergence in political outlook, and (b) *affective polarization*, or a general dislike of political outgroups (Iyengar et al., 2012, 2019; Van Bavel, Rathje, et al., 2021). The latter is reported to have especially increased in recent years, with 41% of Democrats and 45% of Republicans reporting in 2016 that they view the other party as a threat to the nation's well-being, an increase of 10% and 8%, respectively, over 2014, and increasing numbers of Americans reporting discomfort at the prospect of their child marrying someone from

Kimberly Fenn served as action editor.

Jakob Kasper https://orcid.org/0009-0009-1049-7157

None of the data reported in the article have been published elsewhere. The additional online materials are available at https://osf.io/eumt5/?view_only=b2b0b47ea43e43cf9e76e2f3baf8bdb7. The authors declare no competing interests.

Jakob Kasper played a lead role in data curation, formal analysis, and visualization and an equal role in conceptualization, investigation, methodology, project administration, supervision, writing–original draft, and writing–review and editing. Thomas Gilovich played a lead role in funding acquisition and resources and an equal role in conceptualization, investigation, methodology, project administration, supervision, writing–original draft, and writing–review and editing.

Correspondence concerning this article should be addressed to Jakob Kasper, Department of Psychology, Heidelberg University, Hauptstraße 47, 69117 Heidelberg, Germany, or Thomas Gilovich, Department of Psychology, Cornell University, Uris Hall, Ithaca, NY 14853, United States. Email: jakobkasper@icloud.com or tdg1@cornell.edu

the other party (Iyengar et al., 2012; Pew Research Center, 2016, 2022).

The increase in political polarization has been attributed to a number of factors, including structural reforms in the two parties, more fine-grained gerrymandering, the rise of cable news programs and partisan podcasts, and the elimination of the Federal Communications Commission's *fairness doctrine* (Lelkes et al. 2017; Poole, 2008). A great deal of the blame has also been directed at the rise of social media, commonly defined as online platforms that facilitate the production and dissemination of user-created content (Kaplan & Haenlein, 2010). Such platforms are thought to promote polarization by creating homophilic social/political networks that tend to exacerbate myside bias (Del Vicario et al., 2016; Lorenz-Spreen et al., 2020; Van Bavel, Rathje, et al., 2021)—and by triggering greater outrage among users and making it more acceptable for them to express it (Brady et al., 2017; Crockett, 2017; Finkel et al., 2020).

Much of the observed polarization can be understood through the lens of social identity theory, which posits that part of a person's sense of self is derived from the groups to which they belong (Iyengar et al., 2012; Tajfel, 1974; Turner et al., 1987; Van Bavel & Pereira, 2018). Accordingly, a person's self-esteem is tied to the status of their groups, motivating a desire to view their groups favorably and as (positively) distinct from other groups. There are two broad ways of doing so: extolling the ingroup or disparaging outgroups. Although people do both, much of the social identity literature suggests that ingroup favoritism is more common and impactful than outgroup derogation in many domains (Allport, 1954; Balliet et al., 2014; Brewer, 1999; Everett et al., 2015; Hamley et al., 2020; Hewstone et al., 2002; Lee et al., 2022; Lelkes & Westwood, 2017; Rahal et al., 2020). It is often argued that the benefits that come from looking after one's own group are more straightforward (Brewer, 1999, 2007) and that social desirability considerations make it easier to justify beneficial ingroup treatment than harmful outgroup actions (Amira et al., 2021; Hewstone et al., 2002). Consistent with this argument, studies involving non-zero-sum paradigms have shown that people primarily strive to benefit their ingroup (Everett et al., 2015) and pay more attention to the potential outcomes of ingroup members when making decisions (Rahal et al., 2020). Furthermore, studies involving U.S. participants find that most partisans identify more strongly with their own party than "against the opposition" (Lee et al., 2022; Theodoridis, 2019), that affective polarization is more closely tied to bolstering one's party identity than undermining the political outgroup (Amira et al., 2021; Lelkes & Westwood, 2017), and that in-party affinity has a greater impact on voters (Bankert, 2021).

Nevertheless, as the increase in (affective) polarization in the United States suggests (Abramowitz & McCoy, 2019; Abramowitz & Webster, 2016; Finkel et al., 2020), several findings challenge this narrative of ingroup love being more pronounced than outgroup hate. Levels of outgroup derogation tend to be higher when the outgroup is seen as immoral or as a threat to the ingroup's status (Jackson, 1993; Parker & Janoff-Bulman, 2013; Sherif et al., 1954), and cross-party animosity has recently emerged as a strong predictor of certain political beliefs and political participation (Bankert, 2021; Iyengar & Krupenkin, 2018). Therefore, given that both ingroup favoritism and outgroup derogation appear to shape political polarization in certain contexts, researchers have shifted from trying to establish a general preference for one strategy or the other to focusing on context-specific influences on the relative strength of ingroup favoritism versus outgroup derogation (Amira et al., 2021; Bankert, 2021).

Consistent with findings showing a general negativity bias in online communication (Crockett, 2017; Finkel et al., 2020), analyses of Facebook and Twitter posts by politicians and news outlets suggest that social media is a context in which outgroup derogation predominates. For example, mentioning the outgroup has been shown to trigger a greater number of likes/shares than mentioning the ingroup—and to elicit mostly angry reactions in turn (Frimer et al., 2023; Rathje et al., 2021). Outgroup animosity, furthermore, has been shown to be a powerful driver of the sharing of misinformation (Osmundsen et al., 2021), and incivility on social media among political elites has increased by 23% since 2009 (Frimer et al., 2023).

It is unclear from these data, however, whether the preponderance of outgroup derogation over ingroup praise on social media reflects most people's underlying preferences and behavioral tendencies or, instead, something about the structure of the social media landscape. It may be, for example, that most *people* would prefer to share or "like" information that boosts their ingroup while, at the same time, more social media *activity* nevertheless involves attacking outgroups. Indeed, Rathje et al. (2024) found that a representative sample of U.S. participants think that social media posts that criticize the poster's enemies are more likely to go viral—but that they *should not* do so. In the same vein, Heltzel (2019) found that people tend to like those who make an effort to understand the perspectives of their political opponents more than those who decline to do so. Thus, respondents seem unwilling to endorse the kind of invective directed at outgroups that analyses of social media behavior suggest is especially common. What is responsible for this discrepancy between what is found on social media and people's survey responses?

There are multiple factors that might lead to a preponderance of outgroup derogation in analyses of social media behavior, even if the majority of regular users would generally prefer to praise their ingroup (Robertson et al., 2024). First, analyses of social media posts from elite politicians and news broadcasters might reflect the influence of an especially polarized elite and their followers who use negative and moralizing content to mobilize action and maintain political power or media influence (Brewer, 1999, 2007; Iyengar et al., 2012; Rogowski & Sutherland, 2016; Wang & Inbar, 2021; Wojcieszak et al., 2022). It has been shown that although most American Twitter/X users do not follow the accounts of elite politicians, the small group that does has pronounced political biases and is more affectively polarized (Bor & Petersen, 2022; Wojcieszak et al., 2022). For example, a recent analysis of activity on Twitter/X showed that in conversations about national politics among U.S. adults, 6% of the most prolific political tweeters author 73% of the posts (Hughes, 2019). These same users are also more likely to be ideologically extreme and to rate their political outgroup as colder in comparison to other political tweeters, nonpolitical tweeters, or infrequent tweeters. The political discourse online thus seems to be dominated by a small number of individuals with extreme opinions, with neutral or moderate voices being less visible. This increases the volume of divisive content on social media and could have an undue influence on the trends observed in social media analyses (Druckman et al., 2013), while the extreme opinions may not reflect the habits and preferences of most users (Rathje et al., 2024).

In addition to this sort of selection bias, previous research has shown that uncivil content garners more attention (Brady et al., 2017, 2020; Frimer et al., 2023). Algorithms programmed to maximize engagement (e.g., clicks, watch time) may therefore amplify negative content that aligns with an individual's social identity even if that content is not explicitly endorsed by most users. Thus, platform designs may shape behavior by disproportionately exposing users to derogatory messages, leading to higher engagement with this type of content due to factors beyond individual users' control or awareness (Levy, 2021; Van Bavel, Rathje, et al., 2021; Yarchi et al., 2021). Finally, behavioral trends observed on social media platforms can be affected by malign political actors interested in spreading polarizing content via troll farms and bots (Broniatowski et al., 2018). Studies have found that a significant amount of activity on social media is generated by bots (see, e.g., Varol et al., 2017, who estimated that 9%–15% of active Twitter/X accounts were bots), and some are specifically designed to spread polarizing content and sow discord (Broniatowski et al., 2018).

It may be, then, that the average person would prefer to disseminate information that boosts their ingroup even if a majority of the content liked or shared on social media involves outgroup derogation. To determine whether there is a preference for sharing messages that denigrate the outgroup rather than praise the ingroup (Brady et al., 2017, 2020; Crockett, 2017; Finkel et al., 2020), or whether such an apparent preference is partly due to characteristics of the social media environment, it is important to complement the existing research on this question, which relies largely on social media data, with controlled experiments. The following three studies do exactly that. That is, we sought to establish the broader inclinations of U.S. political partisans who are active on social media to gain a more complete understanding of one of the drivers of contemporary political polarization.

Transparency and Openness

For all studies, we report how we determined sample sizes and all data exclusions, manipulations, and measures. Data were analyzed with R (Version 4.2.3, R Core Team, 2023). All three studies were preregistered and approved by the host university's institutional review board. All preregistrations, materials, data, and analysis scripts are publicly available at https://osf.io/eumt5/?view_only=b2b0b47ea43e43cf9e76e2f3baf8bdb7.

Study 1

To examine whether U.S. partisans are more inclined to share social media posts that extoll their ingroup or derogate the outgroup, we recruited U.S. residents who actively use Facebook, Twitter, or both social media platforms and are members of the Republican or Democratic Party. Using a 2 (Intended Impact: Ingroup Favoritism vs. Outgroup Derogation) × 2 (Headline Accuracy: True vs. False) design, we presented participants with news headlines collected from fact-checking websites and asked them whether or not they would share each post. The headlines differed in accuracy (true vs. false), the benefiting party (Republican vs. Democrat), and the intended impact (ingroup favoritism vs. outgroup derogation).

Method

Participants

Participants were recruited from the online platform CloudResearch (Litman et al., 2017) and paid $2.00 for their efforts. We used two filters to restrict the sample to U.S. residents who (a) actively use Facebook and/or Twitter and (b) are members of either the Republican or the Democratic Party ($n_{\text{Republican}} = 100$, $n_{\text{Democratic}} = 100$). The sample size was determined by an a priori power analysis using G*Power (Faul et al., 2009) and based on a conservative estimate of effect size (Cohen's $d = 0.25$), an α level of $\alpha = .05$, and a target statistical power of $1 - \beta = .80$ to test the main hypothesis (i.e., a main effect of intended impact). One participant who provided incomplete data and seven participants who indicated that they did not identify as either Republican or Democratic were excluded, resulting in a final sample of 192 ($M_{\text{age}} = 40.52$, $SD_{\text{age}} = 11.88$); 54.2% of the participants identified as female, 45.3% as male, and one participant did not provide gender information.

Materials

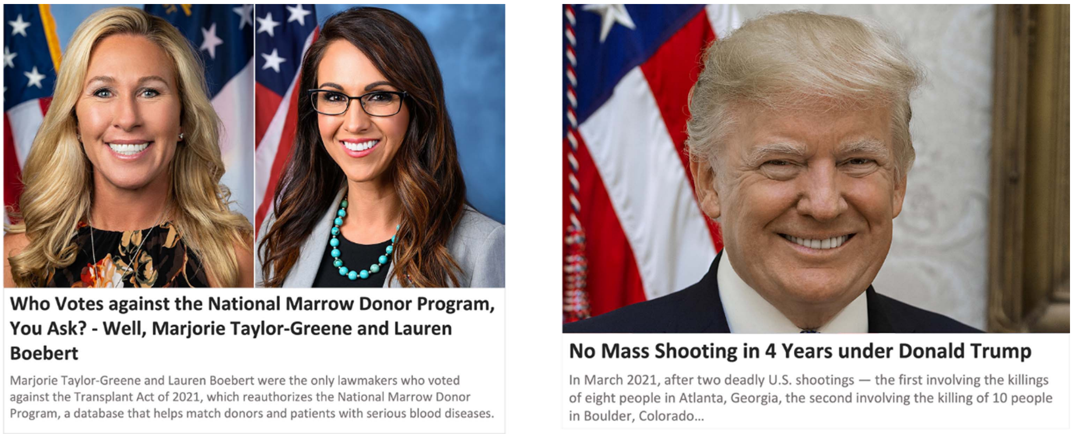
Two sets of true and false news articles (with the accuracy determined by https://www.snopes.com/, https://www.factcheck.org/, and https://www.politifact.com/) were created by collecting political headlines published on these fact-checking websites between January 2021 and March 2022. We first excluded any headline we could not unambiguously classify as serving the interests of the Republican or the Democratic Party. We then further restricted the stimuli to headlines that were classified by the two authors as clear instances of either outgroup derogation (portraying the other party negatively) or ingroup favoritism (portraying Republicans' or Democrats' own party positively). This resulted in five stimuli for each of eight different types of headlines: true/serving Republicans/ingroup favoritism, false/serving Republicans/ingroup favoritism, true/serving Republicans/outgroup derogation, false/serving Republicans/outgroup derogation, true/serving Democrats/ingroup favoritism, false/serving Democrats/ingroup favoritism, true/serving Democrats/outgroup derogation, and false/serving Democrats/outgroup derogation. We used both true and false items for purposes of generality and to explore potential differences between both types of news posts because, for example, misinformation has been shown to be particularly moralizing and negative (Paschen, 2020; Pröllochs et al., 2021; Van Bavel, Harris, et al., 2021) and to spread faster and deeper on social media platforms (Vosoughi et al., 2018).

Each headline was presented with an image, title, and subtitle in a manner designed to resemble articles commonly found on social media (see Figure 1). When constructing the stimuli, the wording of the original headlines was used as much as possible. If the original article itself was not available, we used the information provided by the fact-checking websites to reconstruct the original headline.

To ensure that the headlines that derogate the outgroup and those that extol the ingroup did not differ systematically on other dimensions that might explain our results, we asked an independent

Figure 1

Example Stimuli Used in Studies 1–3



Who Votes against the National Marrow Donor Program, You Ask? - Well, Marjorie Taylor-Greene and Lauren Boebert

Marjorie Taylor-Greene and Lauren Boebert were the only lawmakers who voted against the Transplant Act of 2021, which reauthorizes the National Marrow Donor Program, a database that helps match donors and patients with serious blood diseases.

No Mass Shooting in 4 Years under Donald Trump

In March 2021, after two deadly U.S. shootings — the first involving the killings of eight people in Atlanta, Georgia, the second involving the killing of 10 people in Boulder, Colorado...

Note. The left image is an example of a headline that was classified as true/serving Democrats/outgroup derogation, and the right image is an example of a headline that was classified as false/serving Republicans/ingroup favoritism. The participants saw pictures that were almost identical to the ones shown above. We replaced them with official portraits here for copyright reasons. See the online article for the color version of this figure.

sample of 52 U.S. political partisans ($n_{\text{Republican}} = 26$, $n_{\text{Democrat}} = 26$) who regularly use Twitter and/or Facebook to rate the stimuli that favor their ingroup/derogate their outgroup on the following dimensions: (a) “How compelling, attention-grabbing, or fascinating do you think the headline would be to most people?”; (b) “How emotionally evocative do you think the headline would be to most people—that is, how strong are the emotions the headline would trigger in most people?”; (c) “How prominent, important, and newsworthy is the person who is the subject of the headline?”; (d) “Imagine for a moment that this headline was entirely accurate, how attention-worthy would this news be?” Each rating was made on a scale from *1 = not at all* to *6 = extremely*.

Ratings of how compelling/attention-grabbing/fascinating the content was did not differ between headlines favoring the ingroup ($M = 3.18$, $SD = 1.29$) and headlines derogating the outgroup ($M = 3.19$, $SD = 1.40$), $F(1, 967) = 0.02$, $p = .879$, $R^2_p < .001$. Neither did ratings of how attention-worthy the news would be ($M_{\text{IF}} = 3.20$, $SD_{\text{IF}} = 1.38$; $M_{\text{OD}} = 3.14$, $SD_{\text{OD}} = 1.54$), $F(1, 967) = 0.56$, $p = .456$, $R^2_p = .001$. The headlines extoling the ingroup and derogating the outgroup did differ on emotional evocativeness and newsworthiness, with the subjects in the headlines favoring the ingroup rated as more newsworthy ($M = 4.13$, $SD = 1.21$) than those in the headlines derogating the outgroup ($M = 3.21$, $SD = 1.54$), $F(1, 967) = 146.50$, $p < .001$, $R^2_p = .132$, whereas the headlines derogating the outgroup were rated as more emotionally evocative ($M = 3.14$, $SD = 1.49$) than those favoring the ingroup ($M = 2.93$, $SD = 1.37$), $F(1, 967) = 8.95$, $p = .003$, $R^2_p = .009$. Controlling for these latter two differences, however, did not substantially alter the main results of this study (for more details, see *Appendix Main Analyses Generalized Mixed Models* in the additional online material at https://osf.io/eumt5/?view_only=b2b0b47ea43e43cf9e76e2f3baf8bdb7), with the exception that the significant main effect of intended impact in Part 1, while not substantially differing from the original effect (Gelman & Stern, 2006), was reduced to marginal significance.

Procedure

Participants first provided basic demographic information and then completed three separate parts of the study in which they were presented with the 40 headlines in random order and indicated the following: in Part 1, whether they would share each headline (“Would you consider sharing this story online?”—yes/no); in Part 2, whether they thought each headline was accurate (“To the best of your knowledge, is the claim in the above headline accurate?”—yes/no); and in Part 3, whether they would share each headline knowing its accuracy (“Would you consider sharing this story online, knowing about its accuracy?”—yes/no). In Part 3, participants were told that all headlines had been reviewed by impartial fact-checking organizations and were then shown all the stimuli with each labeled as true or false. The three questions were adopted from Pennycook et al. (2020) and have been shown to predict real-world sharing behavior (Mosleh et al., 2020). We asked first about sharing intentions and then about accuracy (i.e., Parts 1 and 2) because sharing behavior was our primary focus, and it is likely that the measures of participants’ sharing intentions and their accuracy assessments influence each other (Pennycook et al., 2020).

Following Berinsky et al. (2014), we used two types of attention checks. First, there was one additional item in each of the three parts of the experiment that instructed participants to type the word “attention.” Second, after Part 2, participants were asked, “What were your instructions for the second part of the experiment that you just completed (right before this question)?” and had to select the option “Indicate whether the displayed headline is accurate” from three alternatives. All participants remained in the sample for the initial analyses to avoid selective dropout. Replications of analyses without participants who failed one or multiple attention checks are reported in the additional online material at https://osf.io/eumt5/?view_only=b2b0b47ea43e43cf9e76e2f3baf8bdb7 (see *Appendix Analyses Study 1* in the additional online materials).

Results and Discussion

We report mixed logistic regression analyses[1] to model participants' sharing decisions (1 = *yes*, 0 = *no*) and accuracy judgments (1 = *true*, 0 = *false*), crossing headline accuracy (true = 1 vs. false = −1) and headline impact (ingroup favoritism = 1 vs. outgroup derogation = −1). Following the recommendations of Barr et al. (2013), we initially specified models with a maximal random effects structure, including random intercepts for participants and random slopes for accuracy, impact, and their interaction. To avoid issues of singularity and to ensure model parsimony, we conducted likelihood ratio tests comparing the maximal random effects models with models that excluded specific random effects. In this and all subsequent studies, we then selected, for each analysis separately, the simplest model that did not show a significant decrease in fit compared to the maximal random effects model and did not exhibit issues of model convergence or singularity when fitting the model (see *Appendix—Main Analyses Generalized Mixed Models* in the additional online materials at https://osf.io/eumt5/?view_only=b2b0b47ea43e43cf9e76e2f3baf8bdb7 for exact model specifications). Categorization into ingroup favoritism and outgroup derogation was determined by the combination of the stimulus presented (favoring Republicans vs. Democrats) and the participant's party affiliation. We examined the interactions between headline accuracy and headline impact via estimated marginal means (p values Holm–Bonferroni adjusted). Descriptive statistics for Study 1 are presented in Figure 2.

Before participants were informed about the accuracy of each headline, they were more likely to indicate that they would share those that were true, $OR = 1.22$, 95% CI [1.12, 1.33], $p < .001$. More important for present purposes, they were also more likely to report that they would share headlines that favor their ingroup, $OR = 1.27$, 95% CI [1.10, 1.47] $p = .001$. There was no significant interaction between headline accuracy and intended impact ($p = .530$).

In the second phase of the experiment, participants were more likely to judge a headline to be accurate if it was in fact true, $OR = 1.34$, 95% CI [1.24, 1.44], $p < .001$. They were also more likely to indicate that a headline was true if it favored their ingroup, $OR = 1.18$, 95% CI [1.08, 1.30], $p < .001$. These main effects were qualified by a significant interaction between headline accuracy and intended impact, $OR = 1.19$, 95% CI [1.10, 1.28], $p < .001$. For headlines that were in fact true, participants were more likely to consider them true if they favored their ingroup, $OR = 1.97$, 95% CI [1.54, 2.51], $p_{\text{adj}} < .001$. There was no corresponding difference in participants' assessments of the accuracy of false headlines that favored the ingroup or derogated the outgroup, $OR = 0.99$, 95% CI [0.79, 1.25], $p_{\text{adj}} = .945$.

In addition to the direct measure of participants' belief in the claims made in the headlines, we computed a measure of *truth discernment* (the ability to distinguish true and false content) by subtracting the proportion of *yes* responses for the false headlines from the proportion of *yes* responses for the true headlines, separately for content that favored the ingroup and content that derogated the outgroup. Comparing *truth discernment* for content that favors the ingroup and for content that derogates the outgroup with a paired t test, we found that participants were significantly better at distinguishing true from false headlines among those that favored the ingroup ($M = 0.17$, $SD = 0.28$) than among those that derogated the outgroup ($M = 0.05$, $SD = 0.31$), $t(191) = 4.13$, $p < .001$, $d = 0.30$.

After participants were informed about the accuracy of each of the headlines, they stated that they were significantly more likely to share those known to be true, $OR = 31.21$, 95% CI [14.08, 69.14], $p < .001$. Although there was no significant main effect of intended impact, $OR = 0.96$, 95% CI [0.76, 1.21], $p = .709$, there was a significant interaction between intended impact and accuracy, $OR = 1.36$, 95% CI [1.11, 1.67], $p = .003$. Participants were significantly more likely to indicate that they would share true headlines when they favored their ingroup, $OR = 1.69$, 95% CI [1.19, 2.41], $p_{\text{adj}} = .008$, but not significantly more or less likely to indicate that they would share false headlines that favored the ingroup, $OR = 0.50$, 95% CI [0.22, 1.11], $p_{\text{adj}} = .087$.

To summarize, we found that participants have a preference to share information that favors their political ingroup over information that devalues their outgroup.[2] Participants also judged true headlines to be more accurate than false headlines, and their truth discernment was greater for headlines that favored the ingroup. Although participants also tended to think that true headlines favoring their ingroup were more accurate than true headlines derogating their outgroup, this tendency does not account for the observed preference to share information that favors the ingroup more than information that derogates the outgroup—as evidenced by a mediation analysis that found that the direct effect of intended impact on initial sharing decisions remained significant after controlling for its indirect effect via accuracy judgments (see *Appendix Analyses Study 1* in the additional online material at https://osf.io/eumt5/?view_only=b2b0b47ea43e43cf9e76e2f3baf8bdb7). When informed about the accuracy of the statements, participants almost exclusively indicated that they would share accurate content, but they were still more likely to share accurate headlines that favored their ingroup.

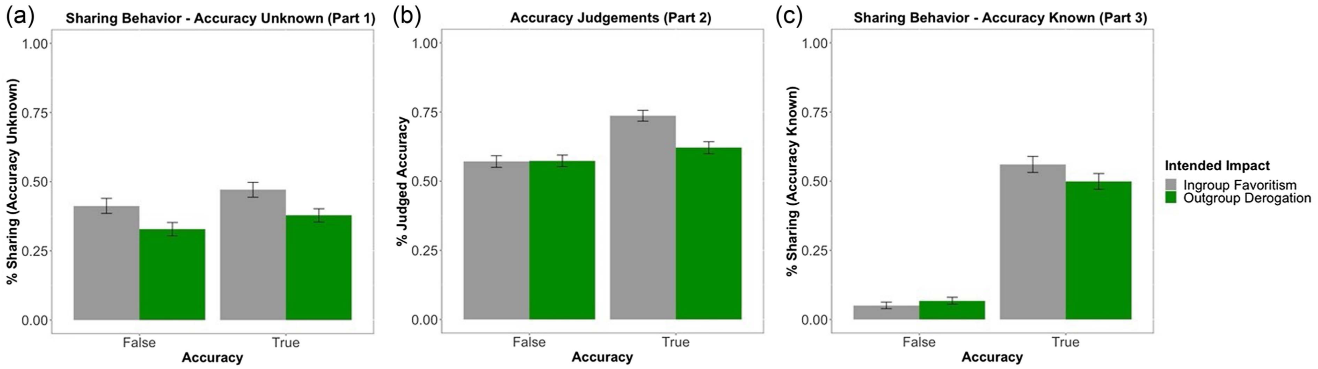
These results suggest a somewhat different picture of the relative appeal of ingroup favoritism versus outgroup derogation than the pattern one could derive from recent analyses of social media activity (Frimer et al., 2023; Osmundsen et al., 2021; Rathje et al., 2021). Although those recent analyses indicate that social media activity that derogates political outgroups is more likely to go viral than activity that boosts users' ingroups, average users may have a general preference to bolster their ingroups.

[1] This is a deviation from our preregistration prompted by our reviewers, who noted the discrepancy between the dichotomous response format (yes vs. no) and the assumptions underlying repeated measures analyses of variance. It is important to note that the results of the two different analyses almost entirely align. The additional online material (https://osf.io/eumt5/?view_only=b2b0b47ea43e43cf9e76e2f3baf8bdb7; see Appendix Analyses Study 1–3) provides the results of the preregistered analyses of variance as well as robustness checks using linear mixed models with crossed random effects for participants and headlines.

[2] When we broke down the results by political affiliation (Republican vs. Democrat), we found that the preference to share information that favors the ingroup over information that disparages the outgroup was significantly stronger for Democratic participants. Because this effect and other main and interaction effects involving ideology were not obtained consistently across our studies and because we were not mainly interested in (and lack the statistical power to reliably detect) further interactions with partisanship, we do not report analyses involving political affiliation here but instead refer the interested reader to the additional online materials at https://osf.io/eumt5/?view_only=b2b0b47ea43e43cf9e76e2f3baf8bdb7 (Appendix—Ideological Differences) that include analyses with ideology as an additional predictor.

Figure 2

Descriptive Statistics for Study 1 (Parts 1–3)



Note. Means and standard errors of the proportion of headlines that participants said they would share when the accuracy of the content was not known (a), when it was known (c), and means and standard errors of the percentage of headlines that participants deemed accurate (b) for Study 1. The accuracy of the headlines (true vs. false) is represented on the *x*-axis. Gray and green bars represent ingroup favoritism and outgroup derogation, respectively. See the online article for the color version of this figure.

Before we can fully trust the reality of the latter effect and confidently conclude that there is a preference to favor the ingroup over derogating the outgroup, we need to examine whether the difference between the present results and those involving analyses of social media data might be due to features of the present study that led participants to express different sharing intentions than they would "in the wild." We conducted the next two studies to test that possibility.

Study 2

Why might we have found evidence of a preference to boost the ingroup when previous analyses of social media activity reported evidence of a preference to derogate the outgroup? One possibility arises from research showing that people are more likely to engage in outgroup derogation when their own group's status and identity are threatened (Amira et al., 2021; Rothschild et al., 2021; Täuber & van Zomeren, 2013). Given that social media is often characterized as rife with hostile content, engaging with social media may make people feel that their group is under threat (Berger & Milkman, 2012; Brady et al., 2020; Crockett, 2017). Perhaps people prefer to extol the ingroup when they do not feel threatened but to derogate outgroups when they do. We tested this interpretation by manipulating whether participants experienced ingroup threat or not.

Method

Participants

Participants were recruited from CloudResearch and paid $2.40 for their efforts. Again, we used two filters to restrict the sample to active social media users who are members of the Republican or the Democratic Party ($n_{\text{Republican}} = 105$, $n_{\text{Democratic}} = 105$). Anyone who took part in the previous study was not eligible to participate. We aimed for a sample size of $N = 200$, relying on an a priori power analysis based on a conservative estimation of the effect size ($f = 0.1$), a target statistical power of $1 - \beta = .80$, and an α level of $\alpha = .05$ to test our main hypothesis (a Condition × Intended Impact interaction). Ten additional participants were recruited to account for potential exclusions. Twelve participants who did not identify with either of the two political parties were excluded, resulting in a sample size of $N = 198$ ($M_{\text{age}} = 41.26$, $SD_{\text{age}} = 13.00$); 55.6% of the participants identified as female and 44.4% as male.

Materials

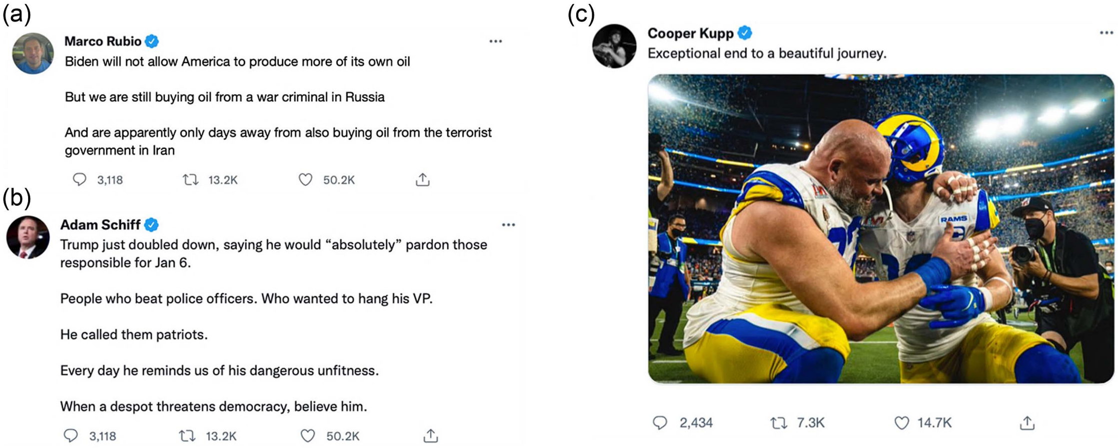
To gather materials that would allow us to manipulate a sense of ingroup threat, we collected tweets posted between January and May 2022 (see Figure 3). We focused on tweets posted by prominent Republicans and Democrats who are active on Twitter. Among those, we specifically selected tweets that attacked the opposing party. We also collected tweets from neutral sources, such as celebrities or sports stars, that had no political content for the control condition. The wording of all tweets was largely in their original form, but with abbreviations written out and links removed. We also removed the date of each tweet and adjusted the number of comments, retweets, and likes to make them approximately equivalent across Republican, Democratic, and neutral sources. Overall, we selected six political tweets (three from Republican and three from Democratic sources), all deemed by the authors to be a clear threat to the opposing party, and six neutral tweets (see *Appendix—Materials and Questionnaires* in the additional online material at https://osf.io/eumt5/?view_only=b2b0b47ea43e43cf9e76e2f3baf8bdb7 for a full list). The news posts used in this study were the same as those in Study 1.

Design and Procedure

The study involved a 2 (Condition: Threat vs. Neutral) × 2 (Headline Accuracy: True vs. False) × 2 (Intended Impact: Ingroup Favoritism vs. Outgroup Derogation) design with condition as a between-subjects factor and headline accuracy and intended impact as within-subject factors. After being randomly assigned to the *threat* or *neutral* condition, all participants were told that in the first

Figure 3

Examples of Stimuli Presented to Participants in Study 2



Marco Rubio
Biden will not allow America to produce more of its own oil

But we are still buying oil from a war criminal in Russia

And are apparently only days away from also buying oil from the terrorist government in Iran

Adam Schiff
Trump just doubled down, saying he would "absolutely" pardon those responsible for Jan 6.

People who beat police officers. Who wanted to hang his VP.

He called them patriots.

Every day he reminds us of his dangerous unfitness.

When a despot threatens democracy, believe him.

Cooper Kupp
Exceptional end to a beautiful journey.

Note. (a) represents a threat to the Democratic Party, (b) represents a threat to the Republican Party, and (c) was one of the neutral tweets. See the online article for the color version of this figure.

phase of the study they would be presented with six randomly selected tweets from a pool of the most popular and impactful tweets in the United States in 2022 (i.e., at least 10,000 likes and 1,000 retweets). Participants in the neutral condition were presented with six neutral tweets, whereas those in the threat condition were presented with three neutral tweets and the subset of three political tweets that attacked their own party, in random order. To increase participants' engagement, they read, for each tweet, "Imagine the author would re-read the tweet a few days after posting it: How satisfied would the author be with the content and the reach of the tweet?" which they answered on a scale from *0* = *not at all* to *6* = *extremely*. At the end of the manipulation phase of the study, participants provided some demographic information (age, gender, social media engagement) and completed two manipulation checks: (a) "I am concerned that information circulating on social media, such as the tweets presented earlier, could tarnish my party's reputation among the general public" and (b) "I am concerned that information circulating on social media, such as the tweets presented earlier, may lead people to question the morality of my party" on a scale from *0* = *not at all* to *6* = *extremely*. The remainder of the experiment was the same as in Study 1, as were data preparation and analyses, but with the additional between-subjects factor of *threat* (*neutral* = −1 vs. *threat* = 1).

Results and Discussion

Attesting to the effectiveness of our manipulation, an independent *t* test revealed that participants in the threat condition reported feeling more threatened ($M = 3.00$, $SD = 2.26$) than those in the neutral condition ($M = 1.67$, $SD = 2.06$), $t(196) = -4.31$, $p < .001$, $d = -0.61$.

A generalized mixed model analysis of participants' initial sharing preferences yielded a significant main effect of intended impact, with participants being more likely to share information that favors their ingroup, $OR = 1.43$, 95% CI [1.25, 1.64], $p < .001$, and a significant main effect of accuracy, with participants being more likely to share factually true headlines, $OR = 1.20$, 95% CI [1.11, 1.30], $p < .001$. There was no significant main effect of *threat*, $OR = 1.11$, 95% CI [0.80, 1.54], $p = .536$, nor were any of the two-way interactions or the three-way interaction significant (see Figure 4).

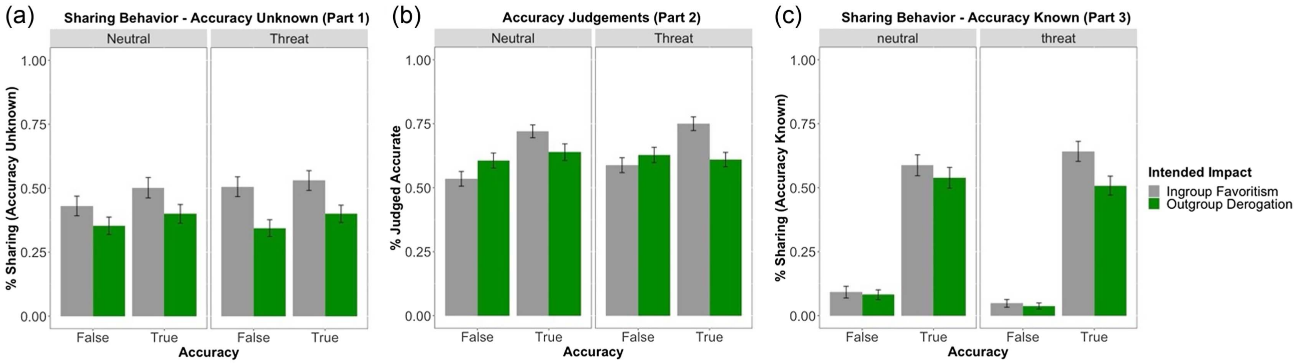
The generalized mixed model analysis of participants' assessments of the accuracy of the news posts yielded a significant main effect of accuracy, with participants being more likely to judge the factually true headlines as accurate, $OR = 1.27$, 95% CI [1.19, 1.37], $p < .001$. There was not a significant main effect of condition, $OR = 1.05$, 95% CI [0.90, 1.22], $p = .558$, but there was a significant main effect of intended impact with participants being somewhat more likely to judge headlines that favor the ingroup as accurate, $OR = 1.09$, 95% CI [1.01, 1.17], $p = .022$. However, a significant interaction between impact and accuracy, $OR = 1.25$, 95% CI [1.16, 1.34], $p < .001$, revealed that for false headlines, participants were *less* likely to judge content favoring the ingroup to be accurate, $OR = 0.76$, 95% CI [0.62, 0.92], $p_{\text{adj}} = .006$, but the opposite was observed for true headlines, with participants being *more* likely to judge content favoring the ingroup to be accurate, $OR = 1.84$, 95% CI [1.50, 2.27], $p_{\text{adj}} < .001$. Neither of the other two-way interactions was significant, nor was the three-way interaction.

A two-way analysis of variance on the truth discernment measure yielded only a main effect of intended impact, with participants being better at distinguishing true and false headlines that favored their ingroup ($M = 0.17$; $SD = 0.28$) than those that derogated the outgroup ($M = 0.01$; $SD = 0.31$), $F(1, 196) = 31.90$, $p < .001$, $R^2_p = .075$.

After being informed about the accuracy of the headlines, participants were much more likely to indicate that they would share true headlines, $OR = 46.55$, 95% CI [20.47, 105.86], $p < .001$, and significantly more likely to share headlines that favored their ingroup, $OR = 1.43$, 95% CI [1.15, 1.78], $p = .001$. There was not a significant main effect of threat, $OR = 0.77$, 95% CI [0.48, 1.23], $p = .274$, and neither of the two-way interactions were significant, nor was the three-way interaction.

Figure 4

Descriptive Statistics for Study 2 (Parts 1–3)



Note. Means and standard errors of the proportion of headlines that participants said they would share when the accuracy of the content was not known (a), when it was known (c), and means and standard errors of the percentage of headlines that participants deemed accurate (b) for Study 2. The accuracy of the headlines (true vs. false) is represented on the *x*-axis. Gray and green bars represent ingroup favoritism and outgroup derogation, respectively. The charts are divided according to the experimental condition (neutral vs. threat). See the online article for the color version of this figure.

To address a colleague's question about whether our findings (and the difference between our findings and previously reported results) might be driven by the responses of participants in our sample who rarely engage with content on social media, we computed the correlation between participants' preference for sharing headlines that favor their ingroup over headlines that derogate the outgroup ($\Delta\text{Sharing} = p[\text{Sharing}_{\text{IF}}] - p[\text{Sharing}_{\text{OD}}]$) and their volume of social media activity, "How often do you share posts on social media (e.g., through Twitter or Facebook)?" as self-rated on a scale from 1 = *never* to 7 = *very often*. Participants' volume of social media activity was positively related to their propensity to share content favoring the ingroup over content disparaging the outgroup in Phase 1 ($r = 0.27$, $p < .001$) but not significantly related to their preference for sharing information that favors their ingroup in Phase 3 ($r = 0.10$, $p = .158$). Clearly, then, our results are not solely due to the preferences of those participants in our sample who are infrequent users of social media (for more details, see *Appendix Analyses Study 2* in the additional online material at https://osf.io/eumt5/?view_only=b2b0b47ea43e43cf9e76e2f3baf8bdb7).

Overall, we largely replicated the results from Study 1, supporting the idea that partisans generally prefer to share content that favors their ingroup over content that derogates the outgroup. Again, the direct effect of intended impact on initial sharing intentions persisted when controlling for its indirect effect via accuracy judgments in a mediation analysis (see *Appendix Analyses Study 2* in the additional online material at https://osf.io/eumt5/?view_only=b2b0b47ea43e43cf9e76e2f3baf8bdb7). Thus, participants' inclination to share headlines favoring their own party more than those denigrating the opposition is not due to the former headlines being seen as more accurate (for the true headlines). This preference was also observed (and descriptively even more pronounced) among participants who were led to feel that their political ingroup was threatened, casting doubt on the possibility that feeling that one's group is under threat is the cause of the preponderance of outgroup derogation observed in previous analyses. Therefore, in the next study, we examine another possible reason for the difference between the results of our first two studies and previous reports in the literature—differences in the motivations of social media users versus participants in laboratory studies. In particular, is it the case that social media users tend to favor content that derogates their outgroups because they think that such content is more likely to "go viral"—a motivation unlikely to be present in laboratory assessments of sharing intentions?

Study 3

Social media platforms have been described as attention economies where clicks/shares are the currency (Lewandowsky & Pomerantsev, 2022). That being the case, social media users' decisions about what to like or share are likely made with an eye toward whether others will like or share in turn (Van Bavel, Rathje, et al., 2021). If people believe that messages derogating the outgroup are more likely to be greeted more favorably and more likely to go viral (see Rathje et al., 2024), that would explain why the users tracked in earlier studies were more interested in sharing such messages than our participants were. We tested this possibility by instructing some participants to make their sharing decisions with an eye toward headlines that are likely to go viral.

Method

Participants

We aimed for a sample size of $N = 200$, based on the same a priori power analysis conducted for Study 2. We recruited 222 active social media users ($n_{\text{Republican}} = 111$, $n_{\text{Democratic}} = 111$) in exchange for $1.90 to meet the target sample size, anticipating some exclusions. Anyone who took part in either of the two previous studies was not eligible to participate. The final sample size was $N = 219$ after excluding three participants without a clear political affiliation ($M_{\text{age}} = 41.38$, $SD_{\text{age}} = 11.89$); 44.7% of the participants identified as male and 55.2% as female.

Materials

To set up our manipulation, all participants read an article about the importance of the upcoming midterm elections with (*viral condition*) or without (*control condition*) a sentence specifically highlighting the parties' desire to spread viral news stories—"modern election campaigns are also carried out on social media platforms *with both parties trying to produce viral posts that are shared by their followers and reach as many people as possible*"(see *Appendix—Materials and Questionnaires* in the additional online material at https://osf.io/eumt5/?view_only=b2b0b47ea43e43cf9e76e2f3baf8bdb7 for the full article). We used the same set of news posts as in the previous studies, but the number of headlines used in the main phase of Study 3 was reduced to 32 (four from each category) to reduce the length of the study.

Design and Procedure

The study consisted of a 2 (Condition: Viral vs. Control) × 2 (Headline Accuracy: True vs. False) × 2 (Intended Impact: Ingroup Favoritism vs. Outgroup Derogation) design with condition as a between-subjects factor and headline accuracy and intended impact as within-subjects factors. Participants were randomly assigned to either the *viral* or *control* condition and received the news article about the upcoming midterm elections with or without (depending on condition) the information about the parties' social media efforts. To determine whether each participant engaged with the material, we asked participants to indicate the topic of the news story after reading the article, and they had to select, from four alternatives, "The importance of the upcoming U.S. midterm elections." The procedures following this initial manipulation differed from those in the previous studies in three respects: (a) As noted above, we used only 32 news posts to ease participants' burden; (b) participants were not asked to assess the accuracy of the news posts (i.e., they did not complete Part 2 of the previous studies) as the sharing intentions were our primary focus and we previously found that intuitions about the accuracy of news headlines could not fully explain the observed sharing preferences; and (c) prior to assessing their initial and fact-checked sharing intentions, participants in the viral condition received the additional instruction: "When considering which headlines to share, please base your decision on a desire to have the news you share go viral."

As a manipulation check, after both Part 1 and Part 2, all participants responded to the question, "In the previous phase of the study, to what extend did you base your decision about what to share on the likelihood that a headline would go viral?" on a scale from 0 = *I did not consider it when making my decisions* to 9 = *it was the sole focus of my decision making*. Furthermore, in addition to basic demographic information (e.g., age, gender), we assessed how much our participants usually engage with content on social media (using the same question as in Study 2) and asked whether they usually share political content ("How often do you share political information on social media [e.g., through Facebook or Twitter]?" on a scale from 1 = *never* to 7 = *very often*). Data preparation and analyses were the same as in Study 2.

Results and Discussion

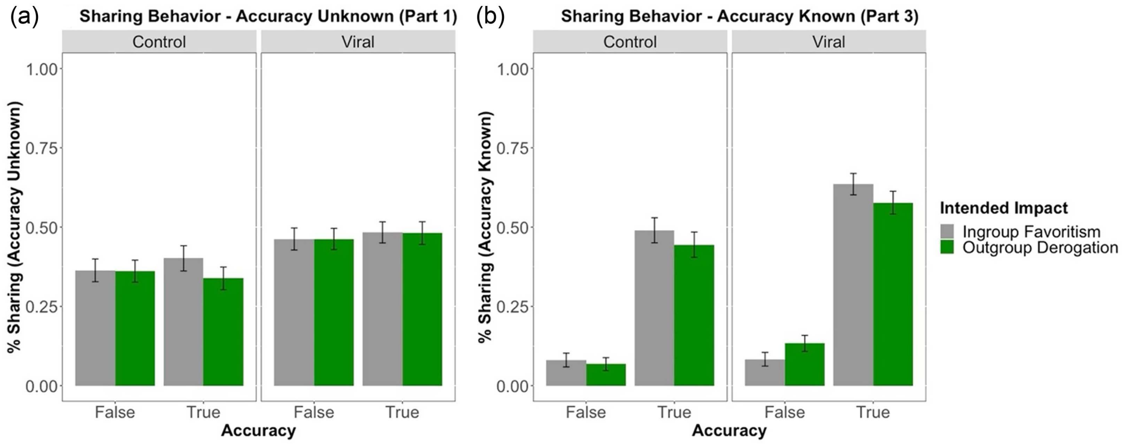
Two independent sample t tests revealed that our manipulation of participants' goal to share viral content was successful, as those in the viral condition were significantly more likely to indicate that they would base their decisions on the likelihood of the stories going viral than participants in the control condition, both when they had not been informed about the accuracy of the headlines ($M_{\text{viral condition}} = 6.87$, $SD = 2.36$; $M_{\text{control condition}} = 1.73$, $SD = 2.64$; $t[217] = -15.20$, $p < .001$, $d = -2.06$), and after they had been informed ($M_{\text{viral condition}} = 6.84$, $SD = 2.26$; $M_{\text{control condition}} = 1.50$, $SD = 2.34$; $t[217] = -17.11$, $p < .001$, $d = -2.32$).

When participants did not know whether the news stories were true, those in the viral condition were more likely to indicate that they would share them, $OR = 1.56$, 95% CI [1.17, 2.08], $p = .002$. No other main or interaction effects were significant (see Figure 5).

In the second phase, when participants knew whether each story was true or not, they were significantly more likely to indicate that

Figure 5

Descriptive Statistics for Study 3 (Part 1 and 3)



Note. Means and standard errors of the proportion of headlines that participants said they would share when the accuracy of the content was not known (a), and when it was known (b) for Study 3. The accuracy of the headlines (true vs. false) is represented on the x-axis. Gray and green bars represent ingroup favoritism and outgroup derogation, respectively. The charts are divided according to the experimental condition (control vs. viral). See the online article for the color version of this figure.

they would share true headlines, OR = 74.26, 95% CI [33.14, 166.40], $p < .001$. There was no significant main effect of condition, OR = 1.57, 95% CI [1.94, 2.60], p = .083, nor was there a significant main effect of intended impact, OR = 0.96, 95% CI [0.77, 1.20], p = .714. However, there was a significant two-way interaction between accuracy and impact, OR = 1.29, 95% CI [1.05, 1.59], p = .014, and a significant three-way interaction between impact, accuracy, and condition, OR = 1.22, 95% CI [1.02, 1.45], p = .026. The two-way interaction reflects the fact that participants were more likely to share headlines that favor the ingroup when they were true, OR = 1.54, 95% CI [1.07, 2.20], p = .039, but not when they were false, OR = 0.55, 95% CI [0.25, 1.20], p = .134. The three-way interaction reflects the fact that the result is further conditional on the experimental condition. More specifically, (a) participants in the viral condition were *less* likely to share false content when it favored the ingroup, OR = 0.26, 95% CI [0.10, 0.66], p_{adj} = .018; and (b) although participants in the viral condition did show a directional preference for sharing true content when it favored the ingroup, OR = 1.58, 95% CI [0.97, 2.59], p_{adj} = .203, and participants in the control condition exhibited such a preference for both true, OR = 1.49, 95% CI [0.87, 2.55], p_{adj} = .287, and false content, OR = 1.18, 95% CI [0.39, 3.59], p_{adj} = .768, none of these latter effects were significant when controlling for multiple comparisons. The other two-way interactions were not significant.

It is important to note, moreover, that neither the extent to which participants usually engage with content on social media (SM1) nor the extent to which they engage with political content (SM2) was significantly correlated with their inclination to share content that favors the ingroup over content that derogates the outgroup (ΔSharing) in Phase 1 ($r_{\text{SM1}} < .01$, p = .956; r_{SM2} = .02, p = .764) or Phase 3 ($r_{\text{SM1}} < .01$, p = .999; r_{SM2} = −.03, p = .641; for more details, see *Appendix Analyses Study 3* in the additional online material at https://osf.io/eumt5/?view_only=b2b0b47ea43e43cf9e76e2f3baf8bdb7 for more details).

Overall, participants in the viral condition were more likely to share content (specifically in Part 1 of the experiment). In contrast to the previous studies, we did not find a significant preference for sharing headlines that favored participants' ingroup. Importantly, we also found no evidence of a preference for outgroup derogation that would mirror previous analyses of social media data, even among participants induced to make their sharing decisions with an eye toward sharing information that was likely to go viral.

Internal Meta-Analysis

We conducted four internal meta-analyses to estimate the average effect of intended impact (ingroup favoritism vs. outgroup derogation) on initial sharing and fact-checked sharing intentions, separately for true and false content. For Studies 2 and 3, we included only participants from the control conditions. All of the analyses used log odds ratios comparing sharing intentions for headlines that favor the ingroup and those that derogate the outgroup. We computed random effects meta-analyses using the Paule–Mandel estimator (Viechtbauer, 2010). The log odds ratios were backtransformed to odds ratios for ease of interpretation. Forest plots displaying the observed outcomes and the estimates based on the random effects models are shown in Figure 6. The p values for analyses focusing on the same part of the experiment (i.e., initial/fact-checked sharing) were Holm–Bonferroni adjusted (for more details, see *Appendix—Internal Meta-Analysis* in the additional online material at https://osf.io/eumt5/?view_only=b2b0b47ea43e43cf9e76e2f3baf8bdb7).

For the initial sharing intentions, although there was considerable variability in the effect sizes across studies, the estimated average odds ratio indicates that participants across the three studies were significantly more likely to say they would be inclined to share a headline if it favored the ingroup than if it derogated the outgroup, both for content that was true, OR = 1.39, 95% CI [1.13, 1.70], z = 3.12, p_{adj} = .004, and content that was false, OR = 1.34, 95% CI [1.00, 1.79], z = 1.97, p_{adj} = .049. For fact-checked sharing intentions, the random effects models indicated that even when participants knew whether the headlines were accurate or not, they were more likely to share headlines that favored their ingroup than those that denigrated their outgroup for true content, OR = 1.32, 95% CI [1.19, 1.47], z = 5.23, $p_{\text{adj}} < .001$, but not for false content, OR = 0.88, 95% CI [0.66, 1.17], z = −0.88, p_{adj} = .376.

General Discussion

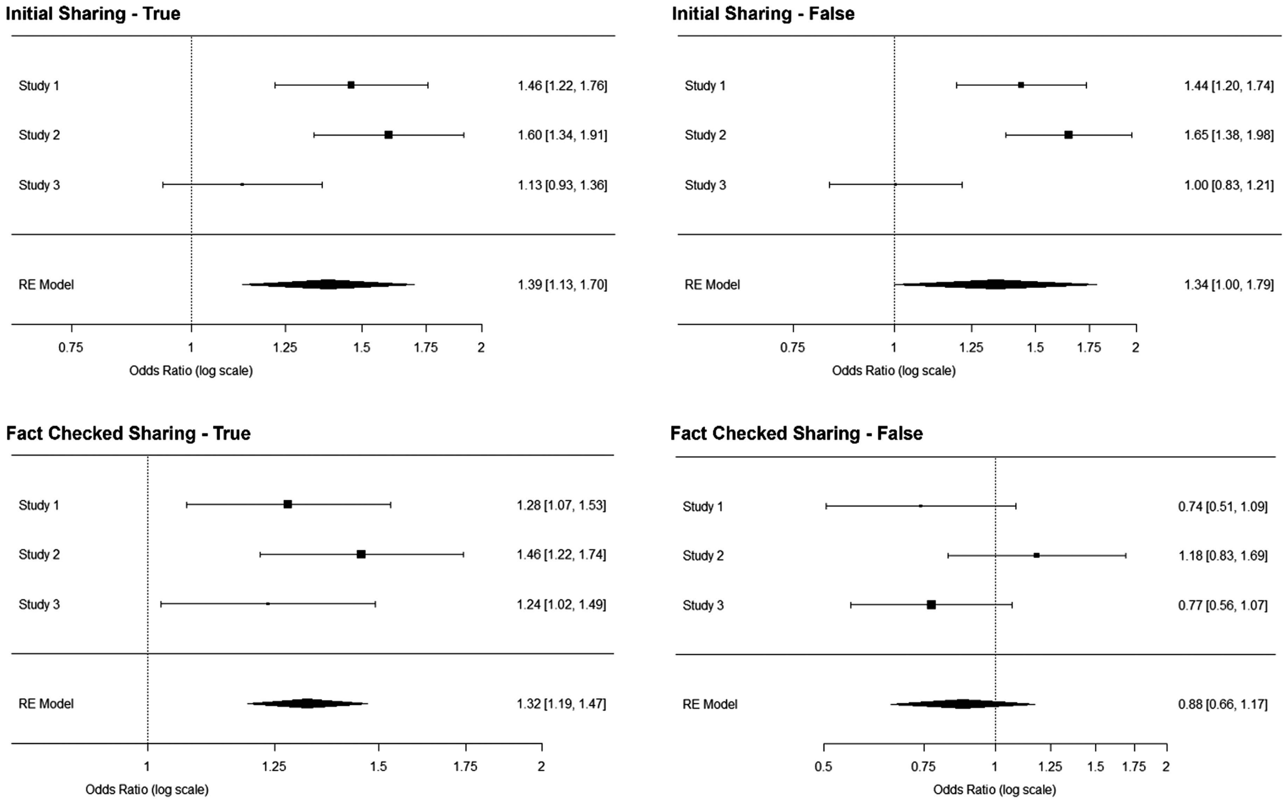
We found that U.S. political partisans prefer to share social media posts that favor their own party rather than denigrate the "other side," a result that expands our understanding of how people think and feel about political content on social media (Heltzel, 2019; Rathje et al., 2024). Our results call attention to the important fact that although social media activity devoted to derogating outgroups garners more likes and shares (Osmundsen et al., 2021; Rathje et al., 2021), that need not entail that the average user has such a preference. Indeed, the average user appears to have the opposite preference. This was particularly pronounced when participants were unaware of whether or not the posts were accurate. When participants *were* aware of whether a headline was accurate or not, the preference for sharing information that favors the ingroup was largely limited to information known to be true. This likely reflects a reluctance to share verifiably false content due to reputational concerns (Altay et al., 2022; Osmundsen et al., 2021).

Although some of our findings indicate that participants also judged (true) headlines favoring their ingroup to be more accurate than headlines derogating their outgroup, their beliefs about the accuracy of the headlines do not seem to provide a comprehensive explanation of the observed sharing preferences. Our mediation analyses established that the direct effect of the intended impact on initial sharing decisions remained stable even after controlling for its indirect effect via accuracy beliefs. Thus, it is unlikely that our participants preferred to share headlines favoring their own party more than those denigrating the opposition simply because the former were thought to be more accurate.

We conducted Studies 2 and 3 to examine whether the difference between the present findings and previous reports of a preponderance of outgroup derogation might be a function of certain features of the real-world social media environment that were not present in our studies. We manipulated a sense of ingroup threat in Study 2 because of research showing that people are more likely to engage in outgroup derogation when their own group's status and identity are under threat (Amira et al., 2021; Rothschild et al., 2021; Täuber & van Zomeren, 2013). Additionally, we addressed participants' social media usage goals in Study 3 because of the common desire on the part of social media users to like or share content that will generate more likes and be further shared (Van Bavel, Rathje, et al., 2021). Neither manipulation,

Figure 6

Results of Internal Meta-Analyses (Studies 1–3)



Note. Observed outcomes and meta-analytical estimates (odds ratios and 95% confidence intervals) for initial sharing intentions (upper panel) and fact-checked sharing intentions (lower panel), separately for headlines that were in fact true versus false. The analyses were carried out using log odds ratios that compare sharing intentions for headlines that favor the ingroup versus those that derogate the outgroup. A RE model was fitted to the data. The log odds ratios were backtransformed to allow for an easier interpretation. RE = random effects.

however, qualified our core finding in a way that would reconcile our results with observed sharing numbers. While we successfully manipulated ingroup threat in Study 2, participants still indicated a preference to share headlines favoring their own party more than those denigrating the opposition—a result that testifies to the robust nature of a preference for ingroup favoritism.

Although focusing participants' attention on virality in Study 3 reduced participants' inclination to share content that favored the ingroup, we did not observe any preference for outgroup derogation in either condition. Thus, even when motivated to post something that will go viral, regular social media users do not appear to prefer to disparage their outgroups rather than praise their own party. The absence of the same significant preference for sharing information that favors the ingroup that we observed in Studies 1 and 2, in the control condition of Study 3, may be due to the fact that we called participants' attention in both conditions of Study 3 to the upcoming 2022 midterm elections. Doing so likely evoked a sense of competition that might have fostered outgroup derogation (Riek et al., 2006; Schlueter & Scheepers, 2010). Still, we also do not find the preference for outgroup derogation that has been observed in analyses of social media behavior.

We readily acknowledge that the sample sizes used in Studies 2 and 3 were not chosen with an eye toward detecting significant (three-way) interactions (as none of our preregistered hypotheses involved such higher-order patterns) and therefore are underpowered when it comes to these questions. We therefore encourage future research that further explores such interactions by utilizing larger sample sizes to ensure greater statistical power.

Constraints on Generality, Limitations, and Future Directions

Our study involved the recruitment of a specific subpopulation of U.S. political partisans who are regular users of social media platforms, and whether our findings can be generalized to other populations or contexts remains unclear. We focus on this subpopulation because it was the focus of earlier research that inspired us to conduct the studies reported here. Future investigations should explore whether the observed preference for sharing content that favors the ingroup can be generalized to other intergroup contexts, such as racial or religious groups, or citizens in countries with different political systems. While there is some evidence of very similar effects being obtained in survey experiments among nationally representative and online samples (Coppock et al., 2018), it would also be desirable to recruit a truly representative sample of U.S. political partisans who use social media. It should be noted, however, that archival analyses of social media data sets can be influenced by small groups of highly active users that, while potentially having significant influences on the online environment and other users, do not necessarily represent the broader population (Rathje et al., 2024; Robertson et al., 2024).

Another question is whether what people *say* they would do in our study corresponds to what they *actually* do on social media. This concern is partly assuaged by (a) research showing that the questions we used to assess participants' sharing decisions reliably predict real-world sharing behavior (Mosleh et al., 2020), (b) by recent research that calls into question evidence of insincere responding in surveys dealing with political issues (Malka & Adelman, 2023), and (c) by our repeated efforts to assure respondents that their responses were anonymous and to stress the importance of responding honestly. It is also important to note that any claim that our results are a misleading product of social desirability concerns rests on the strong assumption that social desirability concerns play out one way online (where attacking the outgroup is seen as socially desirable) and the opposite way in our surveys (where embracing the ingroup is thought to be socially desirable).

Although such a nuanced pattern of social desirability concerns strikes us as implausible (and a decidedly nonparsimonious interpretation of our reported findings), it is of course possible that perceptions of what is socially desirable might differ in the two contexts, thus shaping behavior in different ways. Therefore, to get an empirical handle on whether our results may have been influenced by respondents suppressing their true inclination to derogate the outgroup (and thereby reporting an insincere interest in liking and sharing information that celebrates the ingroup), we analyzed the responses of participants who indicated that they would be willing to knowingly share false content in Part 3 of our three studies—a stated willingness that reflects a notable disinterest in responding in a socially desirable fashion. Did these respondents, who seem to be unconstrained by social desirability concerns, respond in the same way as those who might be more inclined to respond with an eye toward what is socially desirable (i.e., those who never indicated that they would knowingly like or share any false information)?

Yes, they did. In Study 1, the pattern of responses among participants who evidenced little or no concern with social desirability (labeled *SD_Low* below) mirrored those of participants who likely were concerned with what was socially desirable (i.e., those who were unwilling to say they would knowingly spread false information—labeled *control* below).[3] Specifically, in Experiment 1, both subsamples exhibited a preference to share ingroup favoring content in Part 1 of the study, whether that content was true ($M_{\mathrm{IF|SD\_Low}} = 0.70$, $M_{\mathrm{OD|SD\_Low}} = 0.60$; $M_{\mathrm{IF|Control}} = 0.39$, $M_{\mathrm{OD|Control}} = 0.30$) or false ($M_{\mathrm{IF|SD\_Low}} = 0.68$, $M_{\mathrm{OD|SD\_Low}} = 0.58$; $M_{\mathrm{IF|Control}} = 0.32$, $M_{\mathrm{OD|Control}} = 0.24$). The only apparent difference between the two subsamples was that participants in the low social desirability subsample were more inclined to share the headlines they had seen, whether true or false. In the third part of the experiment that examined fact-checked sharing intentions, the observed pattern of responses within both subgroups also aligned for headlines that were true ($M_{\mathrm{IF|SD\_Low}} = 0.77$, $M_{\mathrm{OD|SD\_Low}} = 0.69$; $M_{\mathrm{IF|Control}} = 0.49$, $M_{\mathrm{OD|Control}} = 0.44$). Note that no such comparison is possible for false content because one of the subsamples consisted of individuals who never indicated a willingness to share headlines known to be false. The corresponding data from Studies 2 and 3 also yielded notable alignment between the subsamples presumed to differ in their concern with social desirability (see *Appendix—Social Desirability Check* in the additional online materials at https://osf.io/eumt5/?view_only=b2b0b47ea43e43cf9e76e2f3baf8bdb7 for more details). The congruence between both subgroups across all studies can be taken as evidence against a contaminating effect of social desirability bias on our core findings.

It is also important to stress, once again, that our aim in this research is *not* to refute the findings of previous large-scale analyses of social media data, and thus our findings should not be interpreted as a failed attempt to replicate earlier findings in this area. Rather, our goal is to shed light on a notable gap between observed sharing behavior on social media platforms and surveys of people's sharing preferences. Note that while our studies differ from previous corpus analyses, they are completely in line with what people believe, in the abstract, *should* go viral on social media (Heltzel, 2019; Rathje et al., 2024). For example, Rathje et al. (2024) asked participants to rate on 7-point scales the extent to which social media content in which people criticize their enemies tends to go viral and the extent to which it should go viral. They found that the former was rated significantly above the scale midpoint and the latter significantly below. Our studies build on and extend those findings by showing that such abstract beliefs are also manifested when people make specific, concrete decisions about what to like or share on social media.

This discrepancy between people's preferences and the content most often shared on social media platforms is significant in that it calls attention to the influence of specific features of the social media ecosystem (e.g., algorithms), the potential influence of an especially polarized minority, the difficulties people can have in implementing their behavioral intentions online (i.e., intention behavior gaps), and the different findings that can arise from laboratory and field work—all of which provide fertile ground for future research. It is important to examine further aspects of social media environments and different research methods that may be responsible for the different results that emerge from online observations and experimental methods—and, ideally, to combine both methods in online field experiments (Mosleh et al., 2022).

Finally, because we used only a single manipulation of perceived threat and the desire to go viral (in Studies 2 and 3, respectively), we encourage replications with alternative manipulations. With regards to the threat manipulation, we would like to emphasize that while previous studies mostly utilized threats from neutral sources such as fabricated articles (Amira et al., 2021; Rothschild et al., 2021; Täuber & van Zomeren, 2013), our manipulation was based on a threat common to social media platforms and thereby offers a high degree of ecological validity. Moreover, our manipulation might be expected to be particularly effective in increasing negative partisanship because the source of threat was the relevant outgroup itself. Similarly, explicitly instructing participants to share headlines that are likely to go viral (Study 3) should be seen as a strong manipulation, especially compared to the potentially more implicit drive to generate many likes/shares with each post when scrolling through social media. As pointed out above, future studies should

[3] Note that we are relying on the inspection of descriptive trends in the subgroups due to the small sample sizes in those groups across studies.

aim to manipulate a desire to go viral without conflating it with an intergroup competition framing.

Theoretical and Practical Implications

Why might there be a difference between what is most often shared on social media "in the wild" and what people say they would share when asked under more controlled conditions? That is, what might explain the existence of a relatively aggressive, confrontational social media landscape when, at the same time, individual social media users would prefer to post information that celebrates their ingroup rather than tear down political outgroups? Several possibilities suggest themselves. First, the archival analyses of social media traffic may be influenced by a subset of influential users who are unusually inclined to disseminate derogatory information. Because negative information attracts more attention than positive information (Baumeister et al., 2001; Brady et al., 2020; Rozin & Royzman, 2001), politicians, political activists, news broadcasters, and their followers may engage in outgroup derogation to attract users and boost or maintain their status. Although relatively few in number, the "tracks" they leave on the social media landscape can be pronounced. Second, algorithms may have evolved to take advantage of this very human tendency to be disproportionately drawn to negative information, thus leading to an overrepresentation of derogatory content. Thus, among other features, curated threads may hinder the execution of intentions to promote ingroup favoritism. Finally, the proliferation of content denigrating political outgroups may be due, at least in part, to the influence of malign actors using troll farms and bots to foster societal division by promoting posts that express outgroup animosity. It is therefore important for future research to further examine the impact of these factors in the spread of polarizing content (Alothali et al., 2018; Van Bavel, Rathje, et al., 2021; Yarchi et al., 2021) and to further gauge the amount and type of information distributed by elite party members and their followers (Wojcieszak et al., 2022).

Although our findings do not call into question the prevalence and virality of outgroup derogating messages on various social media platforms, they do support a different image of the average social media *user*, one that, if more widely known, might have beneficial consequences. That is, the belief that people prefer to spread messages that attack various outgroups can, because of its inherent negativity, reinforce societal divisions, discourage constructive discourse, and prompt ever more hostile language online (Finkel et al., 2020; Frimer et al., 2023; Iyengar et al., 2019). Our findings paint a more optimistic picture, one with important practical implications. Correcting the apparent misconception that outgroup hostility is the preferred mode of political discourse on social media may help to lower the temperature of online communication, as negative metastereotypes can lead to increased hatred between parties and reinforce content creators' impulse to put forward hostile information (Lees & Cikara, 2020; Moore-Berg et al., 2020; Robertson et al., 2024). It can also help to overcome social media companies' reluctance to introduce features that would mitigate the spread of divisive content, as their business model is based on maximizing engagement. Our findings thus support the feasibility of altering social media environments to curb the spread of derogatory information—for instance, regulating social media algorithms or counteracting the undue influence of highly polarized elites and their followers (Frimer et al., 2023; Lorenz-Spreen et al., 2020).

References

Abramowitz, A., & McCoy, J. (2019). United States: Racial resentment, negative partisanship, and polarization in Trump's America. *Annals of the American Academy of Political and Social Science*, *681*(1), 137–156. https://doi.org/10.1177/0002716218811309

Abramowitz, A. I., & Webster, S. (2016). The rise of negative partisanship and the nationalization of U.S. elections in the 21st century. *Electoral Studies*, *41*, 12–22. https://doi.org/10.1016/j.electstud.2015.11.001

Allport, G. W. (1954). *The nature of prejudice*. Addison-Wesley.

Alothali, E., Zaki, N., Mohamed, E. A., & Alashwal, H. (2018). Detecting social bots on Twitter: A literature review. *2018 International Conference on Innovations in Information Technology (IIT)* (pp. 175–180). https://doi.org/10.1109/INNOVATIONS.2018.8605995

Altay, S., Hacquin, A.-S., & Mercier, H. (2022). Why do so few people share fake news? It hurts their reputation. *New Media & Society*, *24*(6), 1303–1324. https://doi.org/10.1177/1461444820969893

Amira, K., Wright, J. C., & Goya-Tocchetto, D. (2021). In-group love versus out-group hate: Which is more important to partisans and when? *Political Behavior*, *43*(2), 473–494. https://doi.org/10.1007/s11109-019-09557-6

Balliet, D., Wu, J., & De Dreu, C. K. (2014). Ingroup favoritism in cooperation: A meta-analysis. *Psychological Bulletin*, *140*(6), 1556–1581. https://doi.org/10.1037/a0037737

Bankert, A. (2021). Negative and positive partisanship in the 2016 U.S. presidential elections. *Political Behavior*, *43*(4), 1467–1485. https://doi.org/10.1007/s11109-020-09599-1

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, *5*(4), 323–370. https://doi.org/10.1037/1089-2680.5.4.323

Berger, J., & Milkman, K. L. (2012). What makes online content viral? *Journal of Marketing Research*, *49*(2), 192–205. https://doi.org/10.1509/jmr.10.0353

Berinsky, A. J., Margolis, M. F., & Sances, M. W. (2014). Separating the shirkers from the workers? Making sure respondents pay attention on self-administered surveys. *American Journal of Political Science*, *58*, 739–753. https://doi.org/10.1111/ajps.12081

Bor, A., & Petersen, M. B. (2022). The psychology of online political hostility: A comprehensive, cross-national test of the mismatch hypothesis. *The American Political Science Review*, *116*(1), 1–18. https://doi.org/10.1017/S0003055421000885

Brady, W. J., Gantman, A. P., & Van Bavel, J. J. (2020). Attentional capture helps explain why moral and emotional content go viral. *Journal of Experimental Psychology: General*, *149*(4), 746–756. https://doi.org/10.1037/xge0000673

Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(28), 7313–7318. https://doi.org/10.1073/pnas.1618923114

Brewer, M. B. (1999). The psychology of prejudice: Ingroup love and outgroup hate? *Journal of Social Issues*, *55*(3), 429–444. https://doi.org/10.1111/0022-4537.00126

Brewer, M. B. (2007). The social psychology of intergroup relations: Social categorization, ingroup bias, and outgroup prejudice. In A. W. Kruglanski & E. T. Higgins (Eds.), *Social psychology: Handbook of Basic Principles* (2nd ed., pp. 695–715). Guilford Press.

Broniatowski, D. A., Jamison, A. M., Qi, S., AlKulaib, L., Chen, T., Benton, A., Quinn, S. C., & Dredze, M. (2018). Weaponized health communication: Twitter Bots and Russian trolls amplify the vaccine debate. *American Journal of Public Health*, *108*(10), 1378–1384. https://doi.org/10.2105/AJPH.2018.304567

Coppock, A., Leeper, T. J., & Mullinix, K. J. (2018). Generalizability of heterogeneous treatment effect estimates across samples. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(49), 12441–12446. https://doi.org/10.1073/pnas.1808083115

Crockett, M. J. (2017). Moral outrage in the digital age. *Nature Human Behaviour*, *1*(11), 769–771. https://doi.org/10.1038/s41562-017-0213-3

Del Vicario, M., Vivaldo, G., Bessi, A., Zollo, F., Scala, A., Caldarelli, G., & Quattrociocchi, W. (2016). Echo chambers: Emotional contagion and group polarization on Facebook. *Scientific Reports*, *6*(1), Article 37825. https://doi.org/10.1038/srep37825

Druckman, J. N., Peterson, E., & Slothuus, R. (2013). How elite partisan polarization affects public opinion formation. *American Political Science Review*, *107*(1), 57–79. https://doi.org/10.1017/S0003055412000500

Everett, J. A. C., Faber, N. S., & Crockett, M. (2015). Preferences and beliefs in ingroup favoritism. *Frontiers in Behavioral Neuroscience*, *9*, Article 15. https://doi.org/10.3389/fnbeh.2015.00015

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Finkel, E. J., Bail, C. A., Cikara, M., Ditto, P. H., Iyengar, S., Klar, S., Mason, L., McGrath, M. C., Nyhan, B., Rand, D. G., Skitka, L. J., Tucker, J. A., Van Bavel, J. J., Wang, C. S., & Druckman, J. N. (2020). Political sectarianism in America. *Science*, *370*(6516), 533–536. https://doi.org/10.1126/science.abe1715

Frimer, J. A., Aujla, H., Feinberg, M., Skitka, L. J., Aquino, K., Eichstaedt, J. C., & Willer, R. (2023). Incivility is rising among American politicians on Twitter. *Social Psychological and Personality Science*, *14*(2), 259–269. https://doi.org/10.1177/19485506221083811

Gelman, A., & Stern, H. (2006). The difference between “significant” and “not significant” is not itself statistically significant. *The American Statistician*, *60*(4), 328–331. https://doi.org/10.1198/000313006X152649

Hamley, L., Houkamau, C. A., Osborne, D., Barlow, F. K., & Sibley, C. G. (2020). Ingroup love or outgroup hate (or both)? Mapping distinct bias profiles in the population. *Personality and Social Psychology Bulletin*, *46*(2), 171–188. https://doi.org/10.1177/0146167219845919

Heltzel, G. (2019). *Seek and ye shall be fine: Attitudes towards political perspective-seekers* [Unpublished thesis]. University of British Columbia. https://hdl.handle.net/2429/71391

Hewstone, M., Rubin, M., & Willis, H. (2002). Intergroup bias. *Annual Review of Psychology*, *53*(1), 575–604. https://doi.org/10.1146/annurev.psych.53.100901.135109

Hughes, A. (2019, October 19). *A small group of prolific users account for a majority of political tweets sent by U.S. adults*. Pew Research Center. https://www.pewresearch.org/short-reads/2019/10/23/a-small-group-of-prolific-users-account-for-a-majority-of-political-tweets-sent-by-u-s-adults/

Iyengar, S., & Krupenkin, M. (2018). The strengthening of partisan affect. *Political Psychology*, *39*(Suppl. 1), 201–218. https://doi.org/10.1111/pops.12487

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S. J. (2019). the origins and consequences of affective polarization in the United States. *Annual Review of Political Science*, *22*(1), 129–146. https://doi.org/10.1146/annurev-polisci-051117-073034

Iyengar, S., Sood, G., & Lelkes, Y. (2012). Affect, not ideology: A social identity perspective on polarization. *Public Opinion Quarterly*, *76*(3), 405–431. https://doi.org/10.1093/poq/nfs038

Jackson, J. W. (1993). Realistic group conflict theory: A review and evaluation of the theoretical and empirical literature. *The Psychological Record*, *43*(3), 395–413. https://www.proquest.com/openview/618779f2499c64598670e9f410bd9623/1?pq-origsite=gscholar&cbl=1817765

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, *53*(1), 59–68. https://doi.org/10.1016/j.bushor.2009.09.003

Lee, A. H.-Y., Lelkes, Y., Hawkins, C. B., & Theodoridis, A. G. (2022). Negative partisanship is not more prevalent than positive partisanship. *Nature Human Behaviour*, *6*(7), 951–963. https://doi.org/10.1038/s41562-022-01348-0

Lees, J., & Cikara, M. (2020). Inaccurate group meta-perceptions drive negative out-group attributions in competitive contexts. *Nature Human Behaviour*, *4*(3), 279–286. https://doi.org/10.1038/s41562-019-0766-4

Lelkes, Y., Sood, G., & Iyengar, S. (2017). The hostile audience: The effect of access to broadband internet on partisan affect. *American Journal of Political Science*, *61*(1), 5–20. https://doi.org/10.1111/ajps.12237

Lelkes, Y., & Westwood, S. J. (2017). The limits of partisan prejudice. *The Journal of Politics*, *79*(2), 485–501. https://doi.org/10.1086/688223

Levy, R. (2021). Social media, news consumption, and polarization: Evidence from a field experiment. *The American Economic Review*, *111*(3), 831–870. https://doi.org/10.1257/aer.20191777

Lewandowsky, S., & Pomerantsev, P. (2022). Technology and democracy: A paradox wrapped in a contradiction inside an irony. *Memory, Mind & Media*, *1*, Article e5. https://doi.org/10.1017/mem.2021.7

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods*, *49*(2), 433–442. https://doi.org/10.3758/s13428-016-0727-z

Lorenz-Spreen, P., Lewandowsky, S., Sunstein, C. R., & Hertwig, R. (2020). How behavioural sciences can promote truth, autonomy and democratic discourse online. *Nature Human Behaviour*, *4*(11), 1102–1109. https://doi.org/10.1038/s41562-020-0889-7

Lorenz-Spreen, P., Oswald, L., Lewandowsky, S., & Hertwig, R. (2023). A systematic review of worldwide causal and correlational evidence on digital media and democracy. *Nature Human Behaviour*, *7*(1), 74–101. https://doi.org/10.1038/s41562-022-01460-1

Malka, A., & Adelman, M. (2023). Expressive survey responding: A closer look at the evidence and its implications for American democracy. *Perspectives on Politics*, *21*(4), 1198–1209. https://doi.org/10.1017/S1537592721004096

Moore-Berg, S. L., Ankori-Karlinsky, L.-O., Hameiri, B., & Bruneau, E. (2020). Exaggerated meta-perceptions predict intergroup hostility between American political partisans. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(26), 14864–14872. https://doi.org/10.1073/pnas.2001263117

Mosleh, M., Pennycook, G., & Rand, D. G. (2020). Self-reported willingness to share political news articles in online surveys correlates with actual sharing on Twitter. *PLOS ONE*, *15*(2), Article e0228882. https://doi.org/10.1371/journal.pone.0228882

Mosleh, M., Pennycook, G., & Rand, D. G. (2022). Field experiments on social media. *Current Directions in Psychological Science*, *31*(1), 69–75. https://doi.org/10.1177/09637214211054761

Munn, L. (2021). More than a mob: Parler as preparatory media for the U.S. Capitol storming. *First Monday*, *26*(3). https://doi.org/10.5210/fm.v26i3.11574

Osmundsen, M., Bor, A., Vahlstrup, P. B., Bechmann, A., & Petersen, M. B. (2021). Partisan polarization is the primary psychological motivation behind political fake news sharing on Twitter. *American Political Science Review*, *115*(3), 999–1015. https://doi.org/10.1017/S0003055421000290

Parker, M. T., & Janoff-Bulman, R. (2013). Lessons from morality-based social identity: The power of outgroup “Hate,” not just ingroup “Love.” *Social Justice Research*, *26*(1), 81–96. https://doi.org/10.1007/s11211-012-0175-6

Paschen, J. (2020). Investigating the emotional appeal of fake news using artificial intelligence and human contributions. *Journal of Product and*

Brand Management, *29*(2), 223–233. https://doi.org/10.1108/JPBM-12-2018-2179

Pennycook, G., McPhetres, J., Zhang, Y., Lu, J. G., & Rand, D. G. (2020). Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention. *Psychological Science*, *31*(7), 770–780. https://doi.org/10.1177/0956797620939054

Pew Research Center. (2016). *Partisanship and political animosity in 2016*. https://www.pewresearch.org/politics/2016/06/22/partisanship-and-political-animosity-in-2016/

Pew Research Center. (2022). *As partisan hostility grows, signs of frustration with the two-party system*. https://www.pewresearch.org/politics/2022/08/09/as-partisan-hostility-grows-signs-of-frustration-with-the-two-party-system/?

Poole, K. T. (2008). *The roots of the polarization of modern us politics*. Available at SSRN 1276025. https://doi.org/10.2139/ssrn.1276025

Pröllochs, N., Bär, D., & Feuerriegel, S. (2021). Emotions in online rumor diffusion. *EPJ Data Science*, *10*(1), Article 51. https://doi.org/10.1140/epjds/s13688-021-00307-5

R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rahal, R.-M., Fiedler, S., & De Dreu, C. K. W. (2020). Prosocial preferences condition decision effort and ingroup biased generosity in intergroup decision-making. *Scientific Reports*, *10*(1), Article 10132. https://doi.org/10.1038/s41598-020-64592-2

Rathje, S., Robertson, C., Brady, B., & Van Bavel, J. J. (2024). People think that social media platforms do (but should not) amplify divisive content. *Perspectives on Psychological Science*, *19*(5), 781–795. https://doi.org/10.1177/17456916231190392

Rathje, S., Van Bavel, J. J., & van der Linden, S. (2021). Out-group animosity drives engagement on social media. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(26), Article e2024292118. https://doi.org/10.1073/pnas.2024292118

Riek, B. M., Mania, E. W., & Gaertner, S. L. (2006). Intergroup threat and outgroup attitudes: A meta-analytic review. *Personality and Social Psychology Review*, *10*(4), 336–353. https://doi.org/10.1207/s15327957pspr1004_4

Robertson, C., del Rosario, K., & Van Bavel, J. J. (2024). *Inside the funhouse mirror factory: How social media distorts perceptions of norms*. https://osf.io/kgcrq/download

Rogowski, J. C., & Sutherland, J. L. (2016). How ideology fuels affective polarization. *Political Behavior*, *38*(2), 485–508. https://doi.org/10.1007/s11109-015-9323-7

Rothschild, Z. K., Keefer, L. A., & Hauri, J. (2021). Defensive partisanship? Evidence that in-party scandals increase out-party hostility. *Political Psychology*, *42*(1), 3–21. https://doi.org/10.1111/pops.12680

Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, *5*(4), 296–320. https://doi.org/10.1207/S15327957PSPR0504_2

Schlueter, E., & Scheepers, P. (2010). The relationship between outgroup size and anti-outgroup attitudes: A theoretical synthesis and empirical test of group threat-and intergroup contact theory. *Social Science Research*, *39*(2), 285–295. https://doi.org/10.1016/j.ssresearch.2009.07.006

Sherif, M., Harvey, O. J., Hood, W. R., & Sherif, C. W. (1954). *Intergroup conflict and cooperation: The Robbers Cave experiment*. Wesleyan University Press.

Tajfel, H. (1974). Social identity and intergroup behaviour. *Social Sciences Information. Information Sur les Sciences Sociales*, *13*(2), 65–93. https://doi.org/10.1177/053901847401300204

Täuber, S., & van Zomeren, M. (2013). Outrage towards whom? Threats to moral group status impede striving to improve via out-group-directed outrage. *European Journal of Social Psychology*, *43*(2), 149–159. https://doi.org/10.1002/ejsp.1930

Theodoridis, A. G. (2019). Surprise! Most Republicans and Democrats identify more with their own party than against the other party. *The Washington Post*. https://www.washingtonpost.com/politics/2019/07/25/surprise-most-republicans-democrats-identify-more-with-their-own-party-than-against-other-party/

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory*. Basil Blackwell.

Van Bavel, J. J., Harris, E. A., Pärnamets, P., Rathje, S., Doell, K., & Tucker, J. A. (2021). Political psychology in the digital (mis) information age: A model of news belief and sharing. *Social Issues and Policy Review*, *15*(1), 84–113. https://doi.org/10.1111/sipr.12077

Van Bavel, J. J., & Pereira, A. (2018). The partisan Brain: An identity-based model of political belief. *Trends in Cognitive Sciences*, *22*(3), 213–224. https://doi.org/10.1016/j.tics.2018.01.004

Van Bavel, J. J., Rathje, S., Harris, E., Robertson, C., & Sternisko, A. (2021). How social media shapes polarization. *Trends in Cognitive Sciences*, *25*(11), 913–916. https://doi.org/10.1016/j.tics.2021.07.013

Varol, O., Ferrara, E., Davis, C., Menczer, F., & Flammini, A. (2017). Online human–bot interactions: Detection, estimation, and characterization. *Proceedings of the International AAAI Conference on Web and Social Media*, *11*(1), 280–289. https://doi.org/10.1609/icwsm.v11i1.14871

Viechtbauer, W. (2010). *Conducting meta-analyses in R with the metafor package* (R package Version 2.4-0) [Computer software]. https://doi.org/10.18637/jss.v036.i03

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146–1151. https://doi.org/10.1126/science.aap9559

Wang, S. N., & Inbar, Y. (2021). Moral-language use by U.S. political elites. *Psychological Science*, *32*(1), 14–26. https://doi.org/10.1177/0956797620960397

Whitt, S., Yanus, A. B., McDonald, B., Graeber, J., Setzler, M., Ballingrud, G., & Kifer, M. (2021). Tribalism in AmeFiica: Behavioral experiments on affective polarization in the Trump era. *Journal of Experimental Political Science*, *8*(3), 247–259. https://doi.org/10.1017/XPS.2020.29

Wojcieszak, M., Casas, A., Yu, X., Nagler, J., & Tucker, J. A. (2022). Most users do not follow political elites on Twitter; those who do show overwhelming preferences for ideological congruity. *Science Advances*, *8*(39), Article eabn9418. https://doi.org/10.1126/sciadv.abn9418

Yarchi, M., Baden, C., & Kligler-Vilenchik, N. (2021). Political polarization on the digital sphere: A cross-platform, over-time analysis of interactional, positional, and affective polarization on social media. *Political Communication*, *38*(1–2), 98–139. https://doi.org/10.1080/10584609.2020.1785067

Received June 13, 2023
Revision received July 15, 2024
Accepted September 8, 2024 ■