AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

ISSN: 0096-3445

https://doi.org/10.1037/xge0001374

Tribalism and Tribulations: The Social Costs of Not Sharing Fake News

M. Asher Lawson[1], Shikhar Anand[2], and Hemant Kakkar[3]
[1] Decision Sciences Area, INSEAD
[2] Department of Chemical Engineering, Indian Institute of Technology Delhi
[3] Management and Organizations Academic Area, The Fuqua School of Business, Duke University

Fake news can foster political polarization, foment division between groups, and encourage malicious behavior. Misinformation has cast doubt on the integrity of democratic elections, downplayed the seriousness of COVID-19, and increased vaccine hesitancy. Given the leading role that online groups play in the dissemination of fake news, in this research we examined how group-level factors contribute to sharing misinformation. By unobtrusively tracking interactions among 51,537 Twitter user dyads longitudinally over two time periods ($n = 103{,}074$), we found that group members who did not conform to the behavior of other group members by sharing fake news were subjected to reduced social interaction over time. We augmented this unique, ecologically valid behavioral data with another digital field study ($N = 178{,}411$) and five experiments to disentangle some of the causal mechanisms driving the observed effects. We found that social costs were higher for not sharing fake news versus other content, that specific types of deviant group members faced the greatest social costs, and that social costs explained fake news sharing above and beyond partisan identity and subjective accuracy assessments. Overall, our work illuminates the role of conformity pressure as a critical antecedent of the spread of misinformation.

Public Significance Statement

Fake news contributes to rising political polarization, foments division, and produces contempt for democratic institutions and political outgroups. What group-level factors motivate individuals to share misinformation? By tracking fake news shared online over a 6-month period, we found that people imposed social costs on group members who did not share the same misinformation as them. Social costs were stronger for specifically fake news and among political conservatives relative to liberals. Further experiments revealed that the social costs for failing to share can be as severe as the costs for sharing ideologically opposed content. These results elucidate a worrying mechanism by which group membership encourages the spread of misinformation and impedes the diversity of perspectives available online, providing insight to policymakers considering the regulation of social media and how to combat the rise of fake news.

Keywords: misinformation, group membership, social costs, polarization, fake news

Supplemental materials: https://doi.org/10.1037/xge0001374.supp

Fake news impairs the ability of societies to function (Ecker et al., 2022; van der Linden, 2022). Misinformation has been associated with changes to voting behavior (Allcott & Gentzkow, 2017; Guess et al., 2020; J. Green et al., 2022), reduced compliance with COVID-19 safety measures (Frenkel et al., 2020; Pennycook et al., 2020), and was strongly implicated in contributing to the 2021 insurrection at the United States Capitol (Jervis et al., 2021). The spread of misinformation may undermine trust in political institutions (John, 2021), making it difficult for leaders to govern. This problem became particularly pronounced during the COVID-19

M. Asher Lawson https://orcid.org/0000-0001-5782-3583

The authors declare that they have no competing interests.

All data, code, and materials for Studies 3–6b are available here: https://osf.io/qbkzu/, as are the analysis scripts for Studies 1–2. The raw data for Studies 1–2 cannot be made available because it contains identifiable information from Twitter (usernames, tweet content). M. Asher Lawson presented parts of this work at the 2021 meeting of the International Association for Conflict Management and the 2022 meeting of the Society for Personality and Social Psychology. Hemant Kakkar presented parts of this work at the 2021 Society of Experimental Social Psychology Conference and at the 2021 Society for Personality and Social Psychology Intragroup Processes preconference.

Matthew Asher Lawson served as lead for conceptualization, formal analysis, investigation, methodology, visualization, writing–original draft, and writing–review and editing. Shikhar Anand served as lead for data curation and served in a supporting role for formal analysis and methodology. Hemant Kakkar served as lead for project administration, resources, supervision and contributed equally to writing–review and editing, and served in a supporting role for conceptualization and Methodology.

Correspondence concerning this article should be addressed to M. Asher Lawson, Decision Sciences Area, INSEAD, Boulevard de Constance, 77300 Fontainebleau, France. Email: asher.lawson@insead.edu

pandemic, a period in which compliance with government guidance and mandates became even more important (OECD, 2020). If different groups subscribe to different sources of information during such turbulent periods, rifts in society can emerge as a direct consequence of fake news. Though misinformation is not a novel problem, it has been catapulted to the forefront of social science issues in recent years, partly due to the severity of its consequences.

The rise of social media means that people can widely spread information regardless of their credentials. This can lead to people solely engaging with information that is aligned with their existing beliefs. In general, political echo chambers are said to overexpose people to views similar to their own (Bakshy et al., 2015), which can lead to information asymmetries across groups. These asymmetries can further strengthen people's beliefs in and willingness to share falsehoods that are necessary to maintain their worldviews, which can lead to groups being pushed even further apart. Though research suggests echo chambers may only be a reality for a minority of people (Guess, 2021), they can be harmful nonetheless: for example, in 2016 an armed man traveled to a pizzeria in Washington, DC to break up an alleged child-abuse ring ran by Hilary Clinton after reading about it online (Kang & Goldman, 2016)—an allegation which was baseless. The severe consequences of sharing misinformation prompt the question, why do people share fake news?

Why Do People Share Fake News?

Misinformation can serve a variety of valuable functions for people whether they believe in its truth or not. For example, people may choose to share content that is helpful in supporting their cause. If a news story is consistent with someone's world-view and helps them to protect their social identity, it can be advantageous for them to share it (Kahan, 2017; Schaffner & Luks, 2018). Moreover, sharing fake news can improve the position of one's group, or derogate outgroups (Lawson & Kakkar, 2021; Osmundsen et al., 2021). In these situations, people likely place less value on accuracy (Van Bavel et al., 2021), as news veracity does not matter when the objective is to impress one's outlook on others. Aside from these functional arguments, it is also necessary to consider who is particularly vulnerable to sharing fake news: A wealth of research considers various individual-level predictors of misinformation (e.g., Ecker et al., 2022; Van Bavel et al., 2021; van der Linden, 2022).

For instance, people are more likely to share fake news when they engage in less analytical thinking (Bago et al., 2020; Pennycook & Rand, 2019a, 2019b), have prior exposure to a false story (Pennycook et al., 2018), or do not pay attention to the accuracy of stories (Pennycook et al., 2020, 2021). People are also more likely to believe and share falsities that are consistent with their existing political worldview (Kahan, 2017; Lewandowsky et al., 2012; Van Bavel & Pereira, 2018), and to share fake news as a means to hurt their outgroup due to their political affiliation or a general need for chaos (Abramowitz & Webster, 2016; Lawson & Kakkar, 2021; Osmundsen et al., 2021; Petersen et al., in press). These potent "psychological risk factors," as termed by one recent review (Van Bavel et al., 2021), offer significant progress toward understanding the dynamics of why people share fake news.

Additionally, fake news differs from real news in several meaningful ways that further contribute to its virality. For example, falsities are more likely to evoke emotions such as surprise and disgust, whereas true stories lead to trust and sadness (Vosoughi et al., 2018). Notably, fake news is perceived as more novel than real news (Vosoughi et al., 2018).[1] The emotionality and novelty of fake news could further explain why misinformation travels so fast (Milkman & Berger, 2014). In conjunction with the functional and psychological antecedents of sharing fake news, the properties of the news itself further contribute to its diffusion.

Yet misinformation is inherently a social phenomenon: It is associated with online echo chambers (Cinelli et al., 2021). It is members of the public who disseminate the falsities: they do not circulate on their own (Lewandowsky, 2022). This necessitates considering the group-level and social motives that contribute to people's sharing decisions. Research into group-level motives has shown partisan identity to be a dominant predictor of sharing fake news (Batailler et al., 2022; Osmundsen et al., 2021), and a recent review synthesizes how such group identities can contribute to sharing decisions (Van Bavel et al., 2021). However, existing research does not adequately elucidate how and why group-level factors beyond partisan identity, including the immediate social contexts of group members' sharing decisions, affect individual group members' behavior (Scheufele & Krause, 2019).

Given that people primarily share fake news within online social contexts, it is important to examine the interplay of group-level factors and individuals' social motives in predicting the dissemination of misinformation. Studies examining structural features of people's social networks have found that while endorsing falsehoods may be rare, if this behavior is over-represented in a local neighborhood of a network, it can produce a "majority illusion" (Lerman et al., 2016). This work primarily focuses on modeling how the structural properties of a network can affect people's impressions of the prevalence of beliefs, with related work also highlighting the role of prior exposure and familiarity in fostering belief in fake news (Pennycook et al., 2018). Importantly, rumors and false information tend to circulate more *within* than *across* group boundaries, underscoring the role of group membership in the spread of misinformation (Friggeri et al., 2014). Studies in domains related to misinformation have further demonstrated the role of group membership in endorsing conspiracy theories (Douglas et al., 2017; Ren et al., 2021) and in people exhibiting a need for chaos (Lawson & Kakkar, 2021; Petersen et al., 2018), but there is little understanding of how the social motives associated with group membership drive such behaviors. In the present research, we draw on foundational work in group psychology to understand this phenomenon, contending that group memberships introduce the pressure to conform, which can motivate sharing misinformation.

Group Membership and the Pressure to Conform

Group membership is essential to one's well-being. Individuals derive their sense of self-worth not just from their own characteristics, but also from the groups with which they identify (Tajfel & Turner, 1985). Group membership affords various psychological and social benefits, satisfying individuals' fundamental need to belong (Baumeister & Leary, 1995), providing access to resources

[1] In their seminal study, Vosoughi and colleagues also found that on average false stories reached 1,500 people six times faster than an atypical subset of real stories, and that falsehoods diffused farther, faster, deeper, and more broadly than the truth (Vosoughi et al., 2018), though subsequent work suggests some of these differences are attenuated when controlling for cascade size (Juul & Ugander, 2021).

(Correll & Park, 2005), a sense of collective agency (Bandura, 2000), and reproductive benefits (Caporael, 1997). The belief that a group can both cater to its collective goals and provide strength and affirmation to its members is a core tenet of why group membership is so vital (Bandura, 2000; Correll & Park, 2005). Charles Darwin summarized the functional role of groups by noting that "With those animals which were benefited by living in close association, the individuals which took the greatest pleasure in society would best escape various dangers, whilst those that cared least for their comrades, and lived solitary, would perish in greater numbers" (Darwin, 1896, p. 105).

However, the benefits that group members enjoy often come with explicit or implicit rules that members are expected to follow (Asch, 1956). In other words, an individual member's actions and behaviors are constrained by the norms and procedures of the group. These norms are jointly negotiated rules for social behavior (Sherif, 1936). Specifically, social norms are "rules and standards that are understood by members of a group, and that guide and/or constrain human behavior without the force of laws" (Cialdini & Trost, 1998, p. 152). Norms may emerge as soon as a group is formed, and remain relatively stable after an initial period of flux (Bettenhausen & Murnighan, 1985, 1991).

Failure to conform to group norms can lead to social costs for a group member, such as reduced social interaction or exclusion from the group (Ridgeway & Berger, 1986). Such an experience is highly aversive for most individuals and negatively affects the excluded member's psychological well-being (Williams, 2007). Meanwhile, those who conform are further integrated into groups and become vital figures (Hogg, 2001; Hollander, 1958). Hence, group members are highly motivated to conform in order to avoid social costs. Consistent with this, Centola et al. (2005) showed that fear of punishment can lead to self-enforcement of group norms. In fact, sociological research contends that a defining feature of norms is that individual transgressors are informally sanctioned (Marini, 1984). For instance, group members punish those who violate a group's distributive or cooperative norms (Fehr & Fischbacher, 2004), eschew normative cultural practices (e.g., getting married too young; Settersten & Hagestad, 1996), or violate the sanctity of the group's valued institutions, such as by violating a marriage through seeking a divorce (Liefbroer & Billari, 2010). We draw on this research to suggest that the threat of social costs and resulting conformity pressure affect people's motivation to share fake news.

Social Costs as a Motivator of Sharing Fake News

The research on group conformity and related work on homophily (Asch, 1956; McPherson et al., 2001) discussed so far suggests that group members who conform to the wider behavior of the group will receive social benefits and avoid social costs. We extend this research by contending that similar dynamics will underpin the sharing of fake news. Sharing of misinformation on social media platforms is a visible affair. Someone's online activity is not only visible to those in the individual's social network, but also to the wider public. Online accounts are often publicly open by default (e.g., Twitter), and hence available for anyone to observe a person's online behavior. This makes group members' online activity highly visible, and open to scrutiny and backlash if a member's actions diverge from the group. We thus predict that group members will interact more with group members who share the same fake news as them. Conversely, group members who do not conform by sharing fake news will suffer social costs. Specifically, focal group members who share fake news will reduce their social interaction with deviant members (group members who do not share the same fake news), driving them to the group's margins.

In related work, researchers have found evidence that—rather than truly believing all content that they share—people engage in *expressive responding* to show support for their political side (Schaffner & Luks, 2018). Others suggest that sharing news serves as a form of social authentication, and that social validation is entangled in the processing of news (Waruwu et al., 2021). Taken together, this indicates that sharing fake news that is aligned with one's in-group may be a *primary* means of signaling both political support and group membership. Group members engaging in such participatory propaganda can divert news coverage from other topics that may be harmful to one's political party or ingroup (Lewandowsky et al., 2020) by inducing cascades of supportive information on social media (Lewandowsky, 2022). The act of sharing falsehoods can thus serve a key role in promoting the goals of a group—by influencing public perceptions of it.

Thus, if an individual is aware of the critical social role of sharing falsehoods, and so expects social costs from their group for not sharing misinformation, it is rational for them to share falsehoods to avoid these costs irrespective of their beliefs. There is precedent for this effect in other contexts: For example, group members endorsed honor killings (Vandello & Cohen, 2003), racial segregation (O'Gorman & Garry, 1976), and the communist regime in the former Soviet Union (Kuran, 1995), despite disagreeing with those behaviors privately. In a less extreme setting, both peer and conformity effects are prevalent on online platforms where group members can follow each other's activity. For instance, group members' propensity to "like" Facebook status updates is much higher when the update is endorsed by a majority of users, or when the status update was posted by a friend (Egebark & Ekström, 2011). The threat and severity of such costs are likely to be amplified in homogeneous groups (Huckfeldt et al., 2004), which is of particular concern given the rise of online echo chambers.

Notably, this should not be the case for everyone. Most people will avoid sharing fake news as it could damage their reputation and people's subsequent trust in them (Altay et al., 2022). In fact, over 40% of a study's respondents indicated they would require a payment of $1,000 or more to share fake news (Altay et al., 2022). This is consistent with research demonstrating that the majority of fake news sharing is driven by a minority of prolific sharers (Grinberg et al., 2019). Yet, misinformation is a highly partisan issue (Osmundsen et al., 2021), and sharing fake news that benefits one's group could lead to positive social outcomes in the right context. For example, a Republican sharing news stories that Democrats might call "fake news" could actually enhance that individual's reputation among fellow Republicans, whereas the same content could lead to reputational damage in Democratic circles. In other words, we argue that sharing falsities should lead to positive social outcomes in one's group, as long as the content is endorsed by that group. Reputational *damage* from sharing falsehoods can emerge when such news is not aligned with the group's ideology.

Furthermore, the threat of social sanctions for not sharing content endorsed by the group is one possible contributing factor to the rise and formation of online echo chambers (Bakshy et al., 2015; Cinelli et al., 2021). In such online groups, the selective exposure to perspectives and conformity to other group members' behavior could make individuals' behavior more extreme and resistant to outside perspectives (Bakshy et al., 2015). In fact, in such situations group dynamics are so powerful that individuals may even zealously enforce norms that they do not agree with, just to publicly demonstrate and assert their loyalty to the group (Centola et al., 2005; Willer et al., 2009). If individuals are willing to enforce group norms irrespective of their private beliefs, the likelihood of deviance leading to social costs is even greater. This raises the stakes to comply with the group. Given that these group motives can explain behaviors from homophobia to witch trials (Centola et al., 2005; Erikson, 1966; Willer, 2005), we contend that similar psychological processes will be at play with the sharing of fake news. Group members who do not share the same fake news as others in the group will face social costs in terms of reduced social interaction from those who endorsed such falsehoods.

Though individuals may face costs for deviating from their group's behavior generally, there is reason to believe that failing to share fake news may elicit stronger costs than failing to share real news. Partisan polarization has been discussed as the primary psychological motivator for sharing fake news (Osmundsen et al., 2021). Thus, we should expect those who share fake news to be the most polarized, and these more extreme group members may value in-group loyalty more and be more willing to strongly signal loyalty to the group by imposing social costs. Additionally, it could be the case that sharing fake news is seen as a costly signal—given the potential for reputational damage in some circles as per Altay et al. (2022)—hence, sharing fake news may be a more powerful signal of group loyalty, eliciting stronger social effects. Finally, sharing of fake news is strongly associated with conservative political ideology (Lawson & Kakkar, 2021; Scheufele & Krause, 2019; Vosoughi et al., 2018). Political conservatives may be motivated to endorse false content due to their higher need for shared reality (Jost et al., 2018). We might expect that those who are higher in their need for shared reality would penalize other members harshly who challenge this reality by deviating from the group's shared perspective. In short, we argue that one possible reason why certain groups may share more misinformation is that falsehoods become essential to upholding the group's view of the world, meaning that deviations from such perspectives are unacceptable. For this host of reasons, we predict that social costs will be stronger for failing to share fake news than real news.

Overview of Studies

To study social sanctioning within groups in an ecologically valid manner, it was critical to observe people's naturalistic behavior, free of self-report biases such as social desirability or common method bias. Further, the social contexts that people face in their groups are powerful, and not easily simulated. Hence, to test our social cost hypothesis—that group members reduce their social interaction with those who do not share the same fake news in comparison to those who do—we first collected and analyzed data from Twitter, before conducting further laboratory experiments.

In Study 1, we tracked over 50,000 dyads that were constituent parts of larger groups longitudinally on Twitter from June to December 2020 and analyzed their social interaction patterns after sharing or not sharing the same fake news. Study 2 further augmented this with an additional field study comparing the strength of the relationship between sharing and social interaction in the fake news ecosystem to its strength among a random sample of Twitter users. In two experiments—Studies 3 and 4—we tested causally whether failure to share fake news was associated with reduced social interaction, and if such social costs were greater for not sharing fake news than other content. Study 5 further demonstrated that people take into account social costs in their decisions regarding whether to share fake news, and that social costs explained fake news sharing beyond partisan identity and subjective news accuracy. Finally, Studies 6a and 6b added further nuance, testing the effect of the wider group's behavior on respondents' perceptions of group members who failed to share fake news.

Study Samples and Constraints on Generality

Our overall sample size was large, containing 303,953 observations from 20,546 participants (Table 1). First, we acquired two large field samples from the social media site Twitter (total $n = 281{,}485$ dyads). In our complementary experiments, we also took measures to ensure that our tests were well-powered. We implemented repeated measures designs in all of our studies consisting of many fake news stories to ensure that our results are not an artifact of any single story. We also considered generalizability. Fake news poses a quandary when considering the relevant population of study —1% of individuals account for 80% of fake news source exposures (Grinberg et al., 2019)—which raises the question, should we study these individuals, or the general population? We, therefore, took a balanced approach. For our field Studies 1 and 2, we specifically studied Twitter users who share fake news online. These users varied in the language they used to view Twitter, offering greater diversity than a U.S. only sample, but were not representative of the broader population. Study 2 thus compared these Twitter users to a random sample of Twitter users to help quantify these differences. In our experiments, we used 3 different survey platforms including a Lucid sample that was representative of the U.S. population based on age, gender, ethnicity, and region. The generality of our results is limited to internet users with some connection to the United States—we primarily studied American fake news sources and study participants, and so our results may not generalize in different countries with different languages, media diets, and cultures.

Transparency and Openness

In addition to outlining the constraints on the generality of our conclusions, we also emphasize the steps taken to ensure the transparency and openness of our study designs and analyses. In our OSF repository, we include (a) data for our experiments, (b) analysis scripts that outline the processing of the data, as well as our statistical

Table 1
Sources and Samples for all Studies

Study	Source	Participants	Observations
1	Twitter	12,953	103,074
2	Twitter	4,132	178,411
3	MTurk	500	3,000
4	Lucid	985	3,608
5	MTurk	1,001	10,010
6a	Prolific	488	2,928
6b	Prolific	487	2,922
Total		20,546	303,953

Note. For Studies 1–2, an observation is a dyad; for Studies 3, 4, 6a, and 6b, it is a response regarding a social connection, and for Study 5, it is a response regarding a news story.

analyses, and (c) Qualtrics survey files that show the full details of each study design (Lawson & Kakkar, 2020). For the Twitter studies where it is not possible to make our data public due to privacy concerns, we include our full analysis scripts to show how we reached the results reported in the paper. In addition to these steps, we also include the details of many additional analyses in the online supplemental materials to provide a holistic view of our data.

Study 1

In Study 1, we tested whether a person failing to share a fake or hyperpartisan news story was associated with reduced social interaction from their social connections using observational data from the social media website Twitter. Specifically, we measured the amount of social interaction among a group of users at two-time points (defined as how many public tweets a focal user sent to each social connection), and identified whether these users shared the same fake and hyperpartisan news stories in the period between these two measurements. We tested whether the change in social interaction over time differed for those who shared the same stories relative to those who did not share the same stories. In sum, we tracked social interactions among Twitter users who shared misinformation longitudinally in the period June–December 2020, as well as which specific falsehoods they shared in common.

Twitter Functionality

Twitter is a social network platform where people post public messages, or "tweets." These tweets are viewable on a homepage. Twitter users can choose to "follow" others, forming a unidirectional connection between the accounts. For example, if User A chose to follow User B, User A would now see User B's tweets on their homepage, but not vice versa. Users can post tweets, or can send tweets to other users by mentioning their account username. When a user has posted a tweet, other users can choose to "like" it (signaling approval or interest) or "retweet" it, which will share it with their own follower base. We consider two key types of interactions on Twitter. First, which users share links to fake and hyperpartisan news websites in their tweets, and second, whether the choice to share such content is associated with the amount of public social interaction among Twitter users.

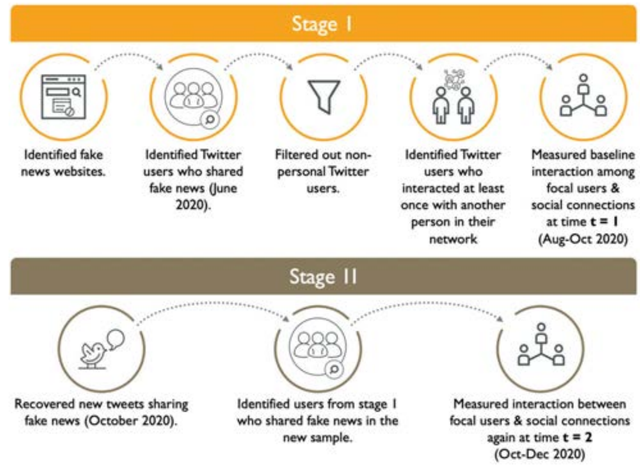
Method

We conducted a two-stage time-lagged study to track how group members' naturalistic social interactions on Twitter changed contingent on whether they shared the same fake news (see Figure 1). In stage 1, we first collected a list of 974 fake and hyperpartisan news websites (see online supplemental materials for further details). Fake websites were those which showed little regard for the truth, whereas hyperpartisan sites had a specific ideological skew without necessarily sharing fake news (Epstein, 2018). We recovered 283,604 tweets sharing links to these fake news websites. These tweets were published by 124,925 unique Twitter accounts. After removing users with more than 5,000 followers or following to focus on individuals who use Twitter for personal reasons (Kivran-Swaine et al., 2011), we were left with a starting sample of 58,872 Twitter accounts.

We distinguished between two kinds of users: focal users (Twitter users who shared fake or hyperpartisan news) and social connections (users in their network with whom they interacted online). As a first

Figure 1

Schematic of Study 1's Data Collection and Processing Procedures



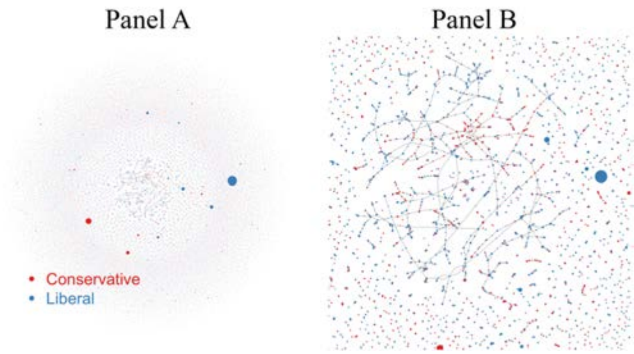
Note. See the online article for the color version of this figure.

step, we measured the baseline social interaction among the focal users and their social connections. We did this by first identifying focal users who had at least one social connection with whom they had a reciprocal conversation (i.e., they exchanged a minimum of one tweet with each other). We then identified all social connections for this set of focal users. Overall, our analysis yielded 21,832 focal users with 161,976 social connections ($M = 7.42$, $SD = 8.04$) resulting in a total sample of 162,663 unique Twitter users.

Our next task was to capture social interaction among these unique group member dyads. We did this by examining directed tweets from the focal users to their social connections. In other words, if User A posted five tweets that mentioned User B directly (i.e., were directed at User B), the social interaction from User A to B would be 5. Separately, if User B posted three tweets that mentioned User A, the social interaction from User B to A would be 3. We chose to measure directed communication rather than the total number of tweets exchanged between the members of each dyad because it enabled us to capture how a social connection's decision of whether or not to share the same story as a focal user determined the tweets that focal user sent to the social connection. It could be the case that a social connection failed to share the same story as a focal user but compensated by sending more tweets to the focal user —which would obscure any social cost coming from the focal user if we did not study *directed* social interaction. The extent of social interaction among these 162,663 unique group members served as a baseline measure for the stage 2 analysis. Figure 2 demonstrates how the interactions between focal users and their social connections were nested in larger social groups.

We began stage 2 data collection by again searching for tweets sharing misinformation using our list of fake news websites (in October 2020). We recovered 272,437 tweets sharing fake stories. We next wanted to identify who among the 162,663 unique stage 1 users shared fake news in stage 2. We, therefore, ran a script with the following logic; (a) identify if a stage 1 user tweeted any fake story, (b) if they did not, drop the user, if they did share, identify their social connections in the dataset, (c) identify what fake stories each social connection shared (if any), and (d) for each connection, measure separately how many fake and hyperpartisan stories they

Figure 2
Social Network Diagram of 42,956 Users in Period t = *1 with Ideology Estimates*



Note. Panel A represents the whole sample, Panel B is an enlarged subset of our data illustrating the densely connected Twitter networks of users sharing fake news. Red (darker) nodes refer to political conservatives, and blue (lighter) nodes political liberals (estimated using a network measure, see Barberá, 2015). The node sizes are scaled by follower count. For further details of the ideology data refer to the Method section. See the online article for the color version of this figure.

shared in common with the focal user. This process allowed us to capture our main binary independent variable: whether each connection shared *any* fake news in common with the focal user, whilst affording greater granularity by measuring the total count of fake news shared jointly by the focal user and social connection, and the type of misinformation shared (fake or hyperpartisan). This resulted in 51,537 dyads of socially interacting users.

Next, we measured our dependent variable—the amount of social interaction within each of the 51,537 group member dyads—for the second period. Again, this measure captured the number of tweets the users publicly sent to each other. There were two measures—the tweets sent from User A to User B, and the tweets sent from User B to User A. This resulted in a total of 103,074 observations of group member dyads' social interactions associated with 12,953 unique focal users.

As an exploratory analysis and to better understand the sample's characteristics, we merged our data with a third-party dataset containing Twitter users' political ideologies recovered using a network-based algorithm (Barberá, 2015). We were able to match 46,666 out of 103,074 observations, with adequate bipartisan representation: the mean ideology (theta) score was 0.049 (slightly right of center), with a broad range (Q1 = −1.27 and Q3 = 1.61). This ideology measure was used in producing Figure 2. We also included several controls to address alternative explanations, such as focal users' network size (total number of followers), the breadth of their information sources (total number of accounts followed), how active they were on the platform (total number of statuses issued and the total number of tweets liked) and the interface language they used (e.g., English, Spanish).

Results

To test whether focal users decreased their social interaction more over time with social connections who failed to share the same fake news as them, we regressed the extent of social interaction initiated by the focal users with a social connection on; whether the social connection shared the same fake news (b_1), time (b_2), and their two-way interaction (b_3), using a random intercept Poisson regression model with observations clustered within each focal user (Table 2). We opted to use Poisson regression because it is efficient in the class of consistent estimators with under or overdispersion (Wooldridge, 2019). The interaction term (b_3) measured the change in social interaction between the focal users and their connections contingent on whether the connections shared at least one of the same fake or hyperpartisan news stories. In Model 1, we found a negative effect of the users sharing the same story ($b = -0.864$, $p < .001$), a negative effect of time period ($b = -2.03$, $p < .001$), and a positive interaction between the two ($b = 1.25$, $p < .001$). In other words, social interaction declined by less if users shared the same fake news, and more if they did not. This interaction remained statistically significant when we included political ideology and the network fixed effects in the model ($b = 1.02$, $p < .001$, Model 2).[2] Concretely, social interaction decreased by an average of 1.77 tweets for social connections who shared the same story but decreased by 1.93 tweets for those who did not share the same story (i.e., there was a social cost of 0.16 tweets).[3] This social cost was 13% of the average degree of social interaction across the periods (1.27 tweets), constituting a sizeable effect.

We also analyzed the data using the *total number* of stories shared in common as an alternative operationalization of our independent variable. The interaction was again statistically significant ($b = 0.370$, $p < .001$, Model 3). Overall, this suggests that not sharing the same fake or hyperpartisan story as the focal users was associated with reduced social interaction from them.

To offer further granularity, we distinguished between users jointly sharing stories from *fake* sources versus *hyperpartisan* sources. The interaction terms between each of these variables (FS and HS) and the time period represented the extent to which focal users changed their interaction with social connections over time as a function of sharing either one additional fake or hyperpartisan news story in common. The interaction between the number of fake stories shared and time period was significant ($b = 0.696$, $p < .001$, Model 4), as was the interaction between the number of hyperpartisan stories shared and time ($b = 0.353$, $p < .001$, Model 4). Notably, the social cost of not sharing was stronger for fake stories than hyperpartisan ones ($\chi^2 = 61.1$,[4] $p < .001$).

[2] This model analyzed the subset of the data for which all of the covariates were available. In the supplementary information, we introduce the political ideology control and network fixed effects in separate models; and provide versions of Models 3–5 without controls, estimated across the whole sample. None of our conclusions are changed by their inclusion or exclusion so we report the models including all controls to be maximally conservative.

[3] Though there is no way to be sure, we believe that social interaction declined on average in our sample for all types of users because we selected users who were socially interacting in the first period; if these users' social interactions were in fact rare, some of them may not have interacted at all in the second period, thus producing this main effect of time. It could also be the case that social interaction within the fake news ecosystem was reduced between these two periods given that between the first and second waves of data collection, uncertainty surrounding COVID-19 was substantially reduced. Neither possible explanation affects our conclusion that interaction reduced *more* for those who did not share the same fake or hyperpartisan stories.

[4] A Wald chi-squared test to compare the fixed effects for a generalized linear mixed-effect model, comparing the size of the coefficients for FS and HS across all of the data (see Table S1 in the supplementary information for further details). In the subset of the data for which we had ideology estimates and other control variables, and when controlling for these additional variables, this difference between the interaction effects shrunk to being marginally significant ($\chi^2 = 3.45$, $p = .067$).

Table 2

Mixed Effects Poisson Regressions: Random Intercepts Clustered at the Focal User Level Predicting Tweets From Focal User to Social Connections

Model	1	2	3	4	5
Intercept	2.558*** (0.012)	2.083*** (0.527)	2.152*** (0.480)	2.149*** (0.482)	2.101*** (0.526)
Connection shared same story (CSSS)	−0.864*** (0.055)	−0.590*** (0.090)			−0.631*** (0.091)
Period (P)	−2.034*** (0.009)	−2.025*** (0.013)	−2.008*** (0.013)	−2.009*** (0.013)	−2.041*** (0.013)
CSSS × P	1.254*** (0.039)	1.021*** (0.065)			1.055*** (0.066)
Political ideology (PI)		0.010 (0.006)	0.009 (0.006)	0.009 (0.006)	0.190*** (0.012)
Number of same shared stories (NSSS)			−0.124* (0.050)		
NSSS × P			0.370*** (0.033)		
Fake stories in common (FS)				−0.245 (0.284)	
Hyperpartisan stories in common (HS)				−0.123* (0.054)	
FS × P				0.696*** (0.178)	
HS × P				0.353*** (0.035)	
SSS × PI					−0.287*** (0.061)
P × PI					−0.163*** (0.009)
SSS × P × PI					0.252*** (0.043)
Control variables	No	Yes	Yes	Yes	Yes
N	103,074	46,353	46,353	46,353	46,353
AIC	294,868	128,008	128,146	128,138	127,657
BIC	294,916	128,375	128,514	128,523	128,051
Log likelihood	−147,429	−63,962	−64,031	−64,025	−63,784

Note. Coefficients for control variables are omitted for simplicity—see SI for full model estimates and additional models.

As an exploratory analysis, we also examined whether liberals and conservatives imposed social costs on their group members to different degrees. We tested the three-way interaction between the binary indicator indicating whether the social connection shared the same fake news story, the time period, and political ideology. We found a significant three-way interaction ($b = 0.252$, $p < .001$, Model 5), showing that the association between not sharing stories and reduced social interaction was *even stronger* for more politically conservative users.

Discussion

Study 1 revealed that social interaction decreased more over time for participants who did not share the same fake or hyperpartisan news stories than for those who did. We hypothesized that this was explained by people choosing to socially interact less with those who do not support them by sharing the same misinformation. Despite the longitudinal nature of our study design, the observed effect cannot be interpreted causally: it could be the case that dyads who were less likely to share the same falsehoods were already drifting in terms of reduced social interaction, or that people were jointly more likely to share the same falsehoods and socially interact more over time, due to some unobserved variable (such as similarity). Hence, despite offering large scale ecologically valid evidence of the link between sharing the same falsehoods and social interaction, further experiments are required to establish whether this relationship is causal.

Study 1 further found that the association between sharing stories in common and social interaction was stronger for specifically fake news relative to hyperpartisan news. This provides tentative evidence that the observed relationship is stronger for fake news than other types of content, but further work is needed to establish the generality of this phenomenon. Given that we selected users from the fake news ecosystem, we were not able to establish whether these effects would also be present among users who share other types of content in common.

Finally, we also observed a three-way interaction between political ideology, sharing stories in common, and time. This effect implies that the association between sharing misinformation in common and social interaction is stronger for more conservative Twitter users. This finding is consistent with recent research demonstrating the greater density in political conservatives' networks on social media than liberals (Chen et al., 2021). A concentrated network facilitates greater visibility of sharing decisions, which could in turn heighten the effect of sharing decisions on social interaction. However, the fake news ecosystem only makes up 0.15% of Americans' daily media diet (Allen et al., 2020) and given that 1% of individuals account for 80% of fake news exposures (Grinberg et al., 2019), this finding may not be representative of typical conservatives and liberals who do not share fake news online.

Study 2

Study 1 found that the failure to share fake and hyperpartisan news was associated with reduced social interaction and that this effect was stronger for fake versus hyperpartisan news. However, the greater social costs for not sharing fake news in comparison to hyperpartisan news were observed within the fake news ecosystem on Twitter (i.e., among users who shared misinformation). It is plausible that similar social costs might be observed in less polarized ecosystems on Twitter where users share any type of content (e.g., celebrity news, real news, other website links). If so, our results would merely show an instantiation of the adverse effects of failure to share any content on social interaction, rather than a *unique* dimension of misinformation that explains its prominence. Thus, in Study 2, we tested for this possibility by examining whether the association between users sharing any content in common[5] and social interaction was stronger in the fake news ecosystem than in a randomly selected sample of Twitter users. Our dataset contained 178,411 new observations of unidirectional social interaction in dyads.

Method

We first randomly sampled 10,000 users from our list of 58,872 Twitter accounts from Study 1 that shared fake news. We recovered all URLs shared by these users in the 9-day window allowed by Twitter's API and excluded users who did not share any URLs. We then measured the social interactions of these users, capturing how many Tweets they sent in the 9-day window and to whom. In Study 2, we focused on social interaction from our focal users to their social connections (i.e., how many tweets User A sent to User B). After removing users who did not share any URLs or send any Tweets in the sample window, we were left with 2,459 focal users who sent 122,091 Tweets to 51,690 unique users. We then searched for all URLs shared by the users that our focal users sent Tweets to. If a link was shared in a retweet, we identified the original Tweet and its associated URL. After compiling the comprehensive list of the URLs shared by the focal users and the users that they tweeted at, we identified how many URLs were shared in common within each dyad. There were 2,141 cases of URLs being shared in common between users. Finally, we retrieved additional fixed effects for the focal users (account language, followers count, friend count, statuses count, and favorites count), in line with Study 1.

To construct a comparison group, we randomly sampled Tweets across all of Twitter for 5 min, removed duplicate users, and randomly sampled 10,000 unique users from the list. We then repeated the steps listed above for this random sample of Twitter users. The final sample of random users contained 2,859 users who sent 530,915 tweets to 150,566 unique users. In this data, there were 17,982 cases of URLs being shared in common. Finally, we merged the data collected from the fake news ecosystem and the random sample of Twitter users and excluded any users with more than 5,000 followers or following, as these likely did not represent personal users of Twitter. The final sample contained 178,411 observations of unidirectional dyadic social interaction in a single time period, associated with 4,132 unique focal users. In other words, the data contained how many tweets 4,132 users sent to each of their social connections (i.e., the level of social interaction), how many stories the focal users shared in common with each of their social connections, and whether the user was from the fake news ecosystem or a random sample of users.

To provide some more intuition regarding the differences in these two groups' media diets, we provide a summary of the most popular sources people shared links to in each group. Specifically, we took the full set of links that each group shared and identified how many links each user shared. For this analysis, we then removed users who shared more than 3 *SD* above the mean number of links to ensure the media diet summary reflected average users rather than these extreme cases. Of the URLs shared by the remaining users, we used a text ,nd to extract all text that came before the first period, and then tabulated the frequencies of each stem and extracted the top twenty entries. The results of this analysis can be seen in Table 3.

The media diets of the two groups are quite different. Though they share some of the same sources (e.g., "dlvr" which is a social media auto-posting platform), the fake news ecosystem's media diet notably includes several fake and hyperpartisan news websites, both on the right (e.g., "thedcpatriot," "order-order," "Newsmax") and the left-wing (e.g., "palmerreport"), as well as conspiracy theory websites (e.g., "dailyexpose"). In sum, the fake news ecosystem sample effectively captured users who frequently shared prominent sources of fake news, and these sources were highly prevalent in the media diet of the group.

Results

We again used hierarchical Poisson regression to predict the total number of tweets a focal user sent to a social connection with the total content shared in common, the type of user being considered, and their two-way interaction. We included random intercepts for each focal user. A positive interaction between the user being from the fake news ecosystem and the amount of content shared in common would indicate that the association between sharing content in common and social interaction was stronger specifically in the fake news ecosystem.

In predicting the degree of social interaction, we found a positive main effect of sharing content in common on the amount of social interaction ($b = 0.020$, $p < .001$) and that members of the fake news ecosystem communicated with their connections less in general ($b = -0.348$, $p < .001$). Crucially, we found a significant positive interaction between shared content in common and the user being from the fake news ecosystem ($b = 0.060$, $p < .001$), both when including controls for characteristics of users' networks (followers, friends, statuses, favorites, and

[5] Content refers to any link to a website (could be fake news, real news, or other content).

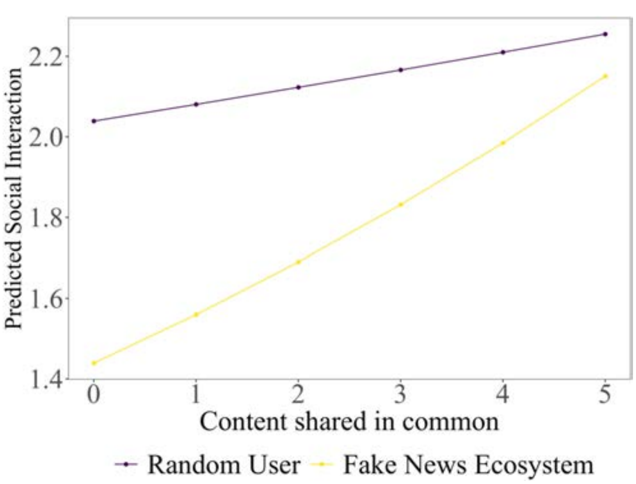
Table 3
The Information Diets of the Real and Fake New Users

Real		Fake	
Source	Links	Source	Links
Naver	4,262	dlvr	1,050
Dlvr	2,522	hill	745
Bit	2,062	a	680
Youtu	1,603	ow	624
blog	759	bit	472
t	729	reut	413
ow	574	thedcpatriot	299
twitch	416	chng	265
urbanopuebla	378	order-order	114
turkcell	365	news	98
opensea	356	palmerreport	89
youtube	316	rapt-plusalpha	85
reut	256	Newsmax	84
listeningparty	244	feedproxy	75
amzn	241	NewsUA	73
blbrd	218	dailyexpose	71
rosea	196	go	70
instagram	193	blog	64
hill	182	disq	62
a	179	richardthekoshimizu	59

language used), as in the reported model, and without. Refer to Table S2 in the online supplemental materials for additional details on these models.

In short, not sharing the same content was associated with lower social interaction in general, but to a stronger degree in the fake news ecosystem. Sharing five pieces of content in common was associated with a 0.22 increase to the number of tweets a focal user sent to a social connection for randomly sampled users, compared to 0.71 tweets for users from the fake news ecosystem (see Figure 3). These increases were 8% and 25.8% of the average total social interaction of 2.75 tweets, respectively. Overall, this interaction effect shows that the relationship between sharing content and social interaction was stronger among users who are prone to sharing fake news.

Figure 3
Predicted Social Interaction Among Users Across Different Levels of Content Shared in Common, Separately for Random Twitter Users and Members of the Fake News Ecosystem



Note. Predictions were made using the model that included control variables (Model 3, Table S2 in the online supplemental materials). The numeric variables were standardized before entering the model and thus were entered as 0 for the predictions. Separate predictions were made for each of the 47 possible user languages and then a single prediction was made by weighting these predictions by the frequency of occurrence of the language across the sample. See the online article for the color version of this figure.

Discussion

Study 2 demonstrated that the association between sharing content in common and social interaction was stronger in the fake news ecosystem relative to a more representative sample of Twitter users. Combined with Study 1's result that the social costs for not sharing specifically fake news were stronger than the social costs for not sharing hyperpartisan news, this suggests that social costs may be greater in the fake news ecosystem.

Importantly, though we observed an association between sharing falsehoods in common and social interaction across Studies 1–2, we cannot offer definite evidence of causality. In Study 1, we measured social interaction longitudinally to mitigate concerns regarding selection effects, but a causal interpretation is still not supported, and in Study 2, we used a cross-sectional design to test for differences in the observed correlation. To support stronger causal inference, we performed five experimental studies. Thus, by using both observational data measuring real-world behavior and experimental data that offers a more controlled environment, we hope to blend external and internal validity to gain a comprehensive understanding of the role of social costs in the spread of misinformation.

Study 3

Study 3 tested whether a social connection's failure to share a news story was causally associated with a reduced desire to socially interact with that connection in a pre-registered experiment. We asked participants to choose six false stories that they would consider sharing, and then evaluate hypothetical friends who did or did not support them by sharing their post. This mimics social media users' experiences, where users see a range of different posts containing news stories, make choices regarding whether to share them, and then decide whether to socially engage with their social connections.

Method

Participants

We recruited 504 participants from Amazon's Mechanical Turk (MTurk) using the CloudResearch platform in exchange for $1.26. Four participants were removed for having non-US IP addresses. We pre-registered our sample size, procedure, and analysis (https://osf.io/dm6je/). The average age of our sample was 40.9 years, and contained 55.4% women, 43.6% men, and 1% people who identify as non-binary. The sample was 63.8% Democrat, 26.4% Republican, and 9.8% Other (generally political independents). We also collected participants' scores on a 10-item Right-Wing Authoritarianism (RWA; Manganelli Rattazzi et al., 2007) scale ($\alpha = .95$). Participants' average scores were left of center (3.08 out of 7), indicating a mild disagreement with the RWA items.

Procedure

After successfully answering a comprehension check question, participants provided informed consent. On the next page, participants faced a list of 12 different partisan fake news stories. Participants selected the three stories that they "would be most likely to share on social media (e.g., Facebook or Twitter)." After selecting these three stories, they were then presented with a list of twelve fake but politically neutral stories to choose from, with the same prompt. Overall, participants chose three partisan and three neutral fake stories (for a total of six stories) that they were most likely to share on social media. Asking participants to select stories they were most likely to share increased the credibility of the follow-up manipulation where they had to indicate their future interaction with a social connection who shared/did not share the same story.

For each story, participants then saw the story headline (in the image-headline format that is common on social media platforms such as Twitter) and were asked to "Imagine you shared this story on your social media." Using a within-subject design for each story, we randomly assigned participants to either a condition where the social connection chose to share, like, or engage with the story that the participant hypothetically shared, or where the social connection chose *not* to share, like, or engage with the story. Participants answered two questions regarding the hypothetical social connection. Participants indicated on 5-point scales the extent to which they were likely to engage in the following behaviors; (a) "Publicly engage with this social media friend by tagging him/her in your post" and (b) "Retweet or share the next post of this friend on the social media." We repeated this procedure for each of the six stories.

After responding to the six stories, participants reported their demographic information (age, gender, education, and race), political ideology (party identification and continuous measure of ideology), and their position on the RWA scale. Participants provided their age in a numeric entry box where responses were allowed between 18 and 120. For gender, participants answered the question, "What gender are you?" with the response options male, female, and non-binary. For political party, participants were asked, "If you had to choose between Democrats and Republicans, who would you prefer?" with the response options Republicans, Democrats, and Other (where "Other" contained a text box to enter a written response). Political ideology was measured on a 7-point scale from "Very liberal" to "Very conservative." Education asked participants to answer, "What is the highest degree that you have earned?" ranging from "Middle School" to "Doctoral," and for race people answered "What is your race or ethnicity?" with the non-exclusive options; white or Caucasian, Hispanic, Black or African American, Native American or Pacific Islander, Asian, and Other. These measures were the same in Studies 3–6b.

Results

Study 3 randomly assigned participants to either a "social connection did share" or "social connection did not share" condition for each story to identify the causal effect of the social connection's behavior on the focal user's desire to socially interact with that social connection. Each of the 500 participants responded to six different news stories for a total of 3,000 observations. Social connections who failed to share the same fake news as the focal user were rated more negatively on both the focal user's desire to publicly interact with them ($M_{\text{Not share}} = 2.58$, $SD_{\text{Not share}} = 1.20$, $M_{\text{Share}} = 2.99$, $SD_{\text{Share}} = 1.21$) and the focal user's likelihood of sharing the social connection's content ($M_{\text{Not share}} = 2.82$, $SD_{\text{Not share}} = 1.12$, $M_{\text{Share}} = 3.17$, $SD_{\text{Share}} = 1.12$). This shows a pattern consistent with the field studies: Individuals were more reluctant to socially interact with social connections who did not share the same stories as them.

To formally test this effect, we estimated regression models predicting each of these two dependent variables (the likelihood of publicly engaging and the likelihood of sharing the social connection's content) using Generalized Estimation Equations, as per our pre-registration.[6] The resulting analysis revealed robust evidence of social penalization across all of our specifications (Models 1–8; all ps $< .001$, Table S4 in the online supplemental materials). As an example model, we predicted respondents' willingness to publicly interact with a social connection with whether the social connection did not share (DNS), the partisanship of the story (PS), and their two-way interaction, when controlling for political ideology, right-wing authoritarianism, education, age, and gender (Model 3, Table S4 in the online supplemental materials). There was a significant negative main effect of DNS ($b = -0.371$, $p < .001$), which constituted the social cost, a significant negative main effect of PS ($b = -0.122$, $p = 0.023$), and a non-significant two-way interaction ($b = -0.036$, $p = .679$). Across models, social connections who did not share the same content as our focal users were evaluated more negatively both in terms of the focal user's desire to publicly engage with those users and in their desire to retweet or share those users' content. Notably, this analysis was robust across the political spectrum. When controlling for both right-wing authoritarianism (Altemeyer, 1981; Manganelli Rattazzi et al., 2007) and political ideology, our results were unaffected.

In addition to finding causal evidence for the effect of failing to share fake news on social interaction, we considered the role of the ideological content of fake news in determining the strength of social costs. To do, we tested whether the effect of a social connection choosing not to share was stronger specifically for Republican-leaning content. Though this interaction was in the direction indicating harsher punishment, it did not attain statistical significance ($b = -0.182$, $p = .118$, Model 4, Table S4 in the online supplemental materials). This differed from the result of Study 1 where social penalization was clearly stronger for more right wing Twitter users. This could be a consequence of a reduced sample size, the different operationalization of ideology, or the different study population.

Discussion

Study 3 found that social connections who did not share the same fake news as study participants were evaluated more negatively—both in terms of desire to publicly engage with those connections and in desire to socially interact with them. This provides causal support for the hypothesis that users face social costs for failing to share fake news stories. However, there are two limitations of Study 3 that necessitate further empirical work: (a) we forced participants to choose stories that they might share, when in reality the true number might be zero, and (b) all of the stories included were fake, which precludes testing whether these social costs varied across the story veracity. We address these two concerns in Study 4.

[6] We used an "independence" correlation structure within cluster to account for the repeated measure structure of our data because this minimized the Quasi-Information Criterion. This is equivalent to linear regression.

Study 4

Study 4 tested whether a social connection failing to share a story in common led to increased social costs. Participants were exposed to a mixture of true and fake stories, asked to indicate any that they would consider sharing, and subsequently answered questions about social connections' behaviors surrounding the stories they indicated they would share. This study allowed us to replicate the results of Study 3 in a new design and test whether social costs varied by news veracity. Notably, whereas in our Twitter studies and Study 3 content was explicitly associated with people's in-groups (e.g., in Study 3, we included partisan content which allowed people to choose content aligned with their political party), this was not the case in Study 4. Respondents were able to choose to share real or fake stories that focused either on COVID-19 or celebrity gossip, but were not explicitly political. We still predict that social costs will be stronger for fake content because for falsehoods that people endorse (by choosing to share them), garnering support from the group is especially important to maintain one's epistemological reality. As noted earlier, the processing of news involves information being authenticated socially (Waruwu et al., 2021): in cases where the news is not true, this social component will be especially important, and thus failure to support a news sharer by spreading their content may lead to harsher social costs from them when the news is fake.

Method

Participants

A priori power analysis indicated that cell sizes of 786 (total $N =$ 3,144) would be sufficient to detect a small interaction effect (0.10 *SD*) with 80% power. We thus recruited a Nationally representative sample of 985 participants from the survey platform, Lucid Theorem (in a repeated measures design we achieved the required cell size to attain sufficient power). Before entering the survey, we verified that participants had US IP addresses and subjected them to a comprehension check. The final sample contained 51.6% women, 47.7% men, and 0.7% non-binary people ($M_{\text{age}} = 45.7$ years). Full details of our pre-registration can be found here: https://osf.io/5s3dk/.

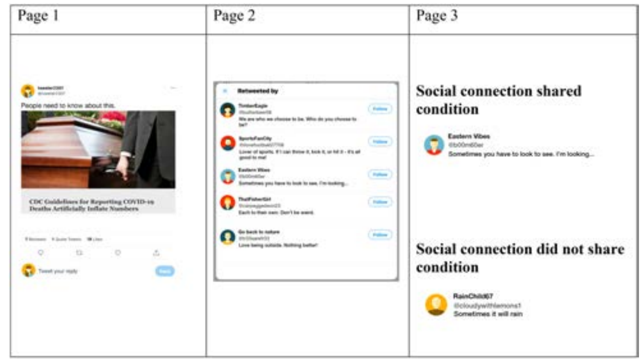
Procedure

Participants were presented with 8 news stories in a random order. Of these stories, 4 were real and 4 were fake, all were politically neutral, and 4 focused on COVID-19 and 4 on celebrity news. For each story, participants were asked, "Would you consider sharing this story online (e.g., through Facebook or Twitter)?" with the response options of "No," "Maybe," and "Yes." For stories where the participant indicated "Maybe" or "Yes" to this question, this was taken to indicate an openness to sharing the story, as per Pennycook et al. (2018) and Pennycook and Rand (2019a). Participants were then shown these stories again and asked questions about a hypothetical social connection.

In the design of this study, we mirrored closely how these stories might appear on social media platforms to avoid any demand effect concerns. Accordingly, for each story that the participant indicated an openness to sharing, they first saw the story again embedded in a graphic designed to mirror a Twitter post (see p. 1, Figure 4). Participants were told, "You said that you would consider sharing the below story on social media. Imagine that you did. Your friends and connections would then decide if they would retweet or favorite your post where you shared the story." On the next page, participants were shown a list of five social connections who chose to retweet the focal post (p. 2, Figure 4) and told, "Imagine some of your friends chose to retweet the story and some didn't. You can see below which of your friends chose to retweet your post. We're next going to ask you some questions about one of your friends." On the final page, participants were again shown the story (and associated Twitter post) and randomly assigned to either consider a social connection who was taken from the list of retweeting users (p. 2) or a different user who did not share the focal story. Participants were shown information about this social connection, including a user icon, display name, handle, and biography.

Figure 4

Study Screens Participants Faced



Note. See the online article for the color version of this figure.

We asked participants to respond to four Likert-scale items regarding the social connection, who either did or did not share the focal story. Specifically, we asked participants on a five-point scale from "Strongly disagree" to "Strongly agree" to "Please indicate your opinion on the statements below. Going forward I would, 'Likely increase my social interaction with this person on social media; Follow the stories shared by this person on social media; Comment on, like, and share the stories posted by this person on social media; Want this person to be in my social network.' We combined these four items to reflect a composite capturing participants" desire to interact with their social connection ($\alpha = .92$). Participants followed this process for each story that they indicated an openness to sharing, before providing demographic information, including age, gender, political ideology, party affiliation, education, and race. These were measured as in Study 3.

Results

To test our hypothesis, we used linear regression with cluster robust standard errors (at the participant level), as per our pre-registration. We controlled for the number of stories shared in total by the participant, and further tested the robustness of our result to including additional fixed effects. With the inclusion of controls for age, gender, political ideology, and education (as well as without), we found a significant positive interaction between the social connection choosing to share a story and the story being fake

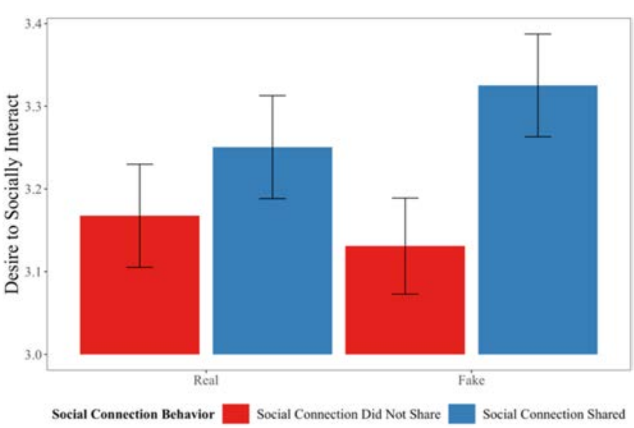
($b = 0.127$, $p = .028$).[7] In this model, the main effects of social connection sharing ($b = 0.061$, $p = .162$) and the story being fake ($b = -0.038$, $p = .291$) were not significant. The number of stories shared by a participant positively predicted their desire to socially interact with all types of social connections, consistent with the idea that such people are more active on social media in general ($b = 0.150$, $p < .001$). Overall, we found that the social costs of not sharing were significantly greater for fake news, corroborating the findings from Studies 1 and 2. This suggests that for content that is important to people maintaining their view of the world, whether that view be associated explicitly with a group such as a political party or content that the sharer views as important to their localized social context, the social costs associated with not sharing will be stronger. Figure 5 presents the average desire to interact contingent on whether the social connection shared or did not share the story and if the stories were real or fake. It appears that there might be a social cost associated with failing to share real news as well, but this effect was not significant.

Discussion

Overall, Study 4 offered further support to our hypothesis. Across Studies 1–4, we consistently found an association between people failing to share falsehoods and their social connections exhibiting a reduced desire to socially interact with them. Studies 1–2 show this association in a correlational manner using ecologically valid observational data, whereas Studies 3–4 replicated the same effect using random assignment to support stronger causal inferences. Studies 1–2 both suggest that the social costs of not sharing were stronger specifically within the fake news ecosystem, and Study 4 replicated this result in a more controlled experiment. We argue that this is because fake news is essential to maintaining the constructed reality of different kinds of groups—so that even when the falsities' content is not explicitly partisan, for those who choose to share it their social connections' support is still essential. We next wanted to test whether individuals consider social costs when making their decisions on whether to share fake news.

Figure 5

Average Desire to Socially Interact by Social Connection Behavior and News Veracity



Note. Error bars indicate 95% confidence intervals. From left to right the bars indicate "Social Connection Did Not Share" for Real News, "Social Connection Shared" for Real News, "Social Connection Did Not Share" for Fake News, and "Social Connection Shared" for Fake News. See the online article for the color version of this figure.

Study 5

In Studies 5–6, we zoom out to consider the wider implications of these results in the context of what we know about the ecosystems that foster misinformation. First, in Study 5, we asked, do people consider such social costs before deciding to share falsehoods? We tested for this possibility, and further examined whether social costs can provide a novel explanation for people's decisions to share above and beyond partisan identity and subjective assessments of accuracy—two well established predictors of fake news behavior (Osmundsen et al., 2021; Pennycook et al., 2021).

Method

Participants

One thousand one MTurk participants completed the study in return for \$0.71 (0.8% non-binary, 53.7% women, 45.5% men, $M_{age} = 42.2$ y), after being recruited using CloudResearch.[8] Participants reported their identification with the Democrat or Republican party on a 6-point scale ranging from Very Strong Democrat to Very Strong Republican, including a separate option for "Neither." 141 participants reported they identified with neither party.

Procedure

Participants viewed 10 different fake news stories in a random order, of which 5 were associated with COVID-19 and 5 were associated with celebrities (total observations = 10,010). After viewing each story, participants were first asked to answer, "To the best of your knowledge, how accurate is the claim in the above headline?" on a 4-point scale from "Not at all accurate" to "Very accurate." We counterbalanced the order of the next two sets of questions—participants either answered whether they would share the story first, before next reporting on their beliefs about their social connections, or vice versa.

For the sharing item, participants were asked "Would you consider sharing this story online (e.g., through Facebook or Twitter)?" with the possible responses of "No," "Maybe," and "Yes." Both "Maybe" and "Yes" were coded as a 1, signifying openness to sharing a story. For the social items participants were asked to "Please indicate your agreement with each of the following statements on a 5-point scale from 'Strongly disagree' to 'Strongly agree'". Your social connections would "Interact with you more if you shared this story; Appreciate you sharing this story; Interact with you less if you didn't share this story; Leave you out of interactions if you didn't share this story." Without any reverse coding, we combined these four items into a single

[7] We look at the benefit of social connections sharing content in common (as in Study 2), rather than costs associated with not sharing.

[8] We did not apply a bot detection procedure here. In Study 6b, our bot checking procedure identified 9/493 respondents were potential bots in a similar MTurk sample. If the rate were similar in this study, it would be unlikely to change any of our focal effects (all $p < .001$).

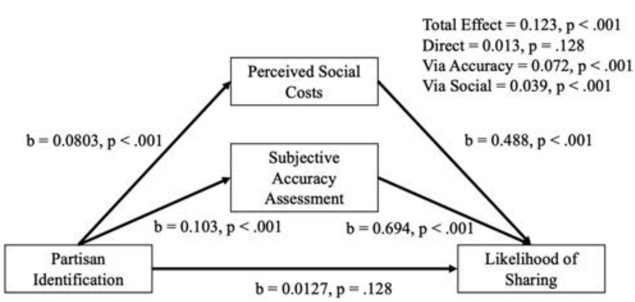
measure of social costs that had an alpha of 0.84. Finally, demographic information (age, gender, political ideology, partisan identification, education, and race) was collected with the same procedure as Studies 3 and 4.

Results

As a first step, we wanted to establish whether social costs provided an explanation for why people share fake news above and beyond explanations based on subjective accuracy beliefs or partisan identity. To do so, we first excluded participants who responded “Neither” on our partisan identity measure (we analyze these data in the online supplemental materials). Using logistic regression with cluster robust standard errors to predict participants’ decisions on whether they would share a story, we replicated the effect of partisan identity and accuracy on sharing decisions. Partisan identity ($b = 0.097$, $p = .011$) and subjective accuracy assessments ($b = 1.58$, $p < .001$) both significantly predicted the likelihood of a participant choosing to share a story. When including our social cost variable, the model fit improved ($R^2 = .371$ to $R^2 = .433$). Using an ANOVA test, we found that the inclusion of the social cost variable contributed significant explanatory power ($\chi^2 = 194.5$, $p < .001$). The effect of social costs on sharing was positive and statistically significant ($b = 0.777$, $p < .001$) in this model, when including partisan identity ($b = 0.070$, $p = 0.053$), which became marginally significant, and subjective accuracy assessments ($b = 1.49$, $p < .001$) in the regression equation. In other words, when participants perceived that there were greater social costs associated with not sharing a story, they were more likely to share that story. This effect was distinct from the effect of partisan identity on sharing decisions.

To further examine if social costs might *explain* the relationship between partisanship and sharing fake news beyond explanations based on subjective accuracy, we estimated a structural model including social costs and subjective accuracy as two parallel mediators (Figure 6). Upon inclusion of the two mediators, the direct effect of partisan identification on sharing became insignificant (suggesting full mediation), but there were significant indirect effects of partisan identification both via social costs ($b = 0.039$, $p < .001$) and accuracy ($b = 0.072$, $p < .001$). Seeking to clarify the role of social costs above and beyond partisanship, we swapped partisan identity and social costs around in the model, finding that the direct effect of social costs on sharing was greater than the indirect effect via partisan identity ($z = 17.2$, $p < .001$). Overall, this analysis revealed social costs as a distinct psychological pathway explaining the sharing of fake news beyond partisan identity and accuracy beliefs.

Figure 6
Model With Perceived Social Costs and Subjective Accuracy Assessments as Parallel Mediators



Finally, we analyzed the four-item social cost scale as two subdimensions of two items each, focused on social rewards and punishment, respectively. This was done to examine whether anticipation of social rewards (i.e., greater interaction with group members after sharing) or fear of social penalties (i.e., reduced interaction or exclusion for not sharing) was driving the effect, as social costs could reflect both reduced social rewards and greater social penalties. We observed significant indirect effects of partisan identification via both pathways ($p < .001$, Table S10 and Figure S2 in the online supplemental materials), suggesting support for both anticipation of social rewards and fear of social penalties were viable underlying mechanisms that might drive sharing behavior.

Discussion

Overall, Study 5 revealed that social costs were a unique predictor of sharing fake news above and beyond existing explanations based on partisanship and accuracy beliefs. These factors are of course correlated—for example, partisanship might elevate both subjective accuracy beliefs and perceived social costs, and social costs might influence people’s accuracy ratings if they internalize these pressures. Study 5’s data further suggested that social costs were a key pathway explaining the link between partisan identity and sharing fake news. Overall, the evidence suggests that social costs are an important motivator of people’s decisions to share misinformation. In Studies 6a and 6b, we manipulated the behavior of respondents’ social networks to offer greater nuance in understanding the precise nature of our effects.

Study 6a

Study 6a examined how respondents evaluated deviant group members. Whereas Studies 3–4 focused on evaluations of a single social connection who deviated from the survey respondent’s choices, Study 6a presented the behavior of a hypothetical group and asked participants to evaluate members of them. We measured participants’ evaluations of group members’ accuracy and their desire to socially interact with them, allowing us to decompose the nature of our effect further.

Method

Participants

We recruited a sample of 490 participants on the online platform Prolific, who were paid $1.41 for their participation. As per our preregistration, two participants were removed after collecting data for having a duplicate or non-US IP address (https://osf.io/sutgj/).[9] The sample was 52.3% men, 2.3% non-binary, and 45.5% women, with an average age of 35.3 years. We again observed a reasonable number of people who did not identify as either Republican

[9] We included a screen that identified potential bots by flagging users who clicked fewer times than the number of response boxes on a page, but it only identified 1/488 participants as a potential bot, so we completed all of our analyses with the full sample.

or Democrat—our sample was 15.2% Republican, 73.8% Democrat, and 11.1% neither Republican nor Democrat.

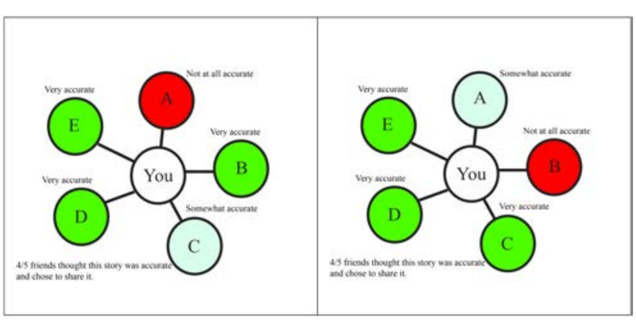
Procedure

Participants faced six different fake news stories. Of these stories, three were Republican-leaning and three were Democrat-leaning. These stories were randomly drawn from a total pool of 12 items (of which six were Republican-leaning and six were Democrat-leaning). First, participants read the news story with a brief headline and picture. Following this, participants learned how their immediate hypothetical social network responded to the story. In all cases, four out of five nodes of the hypothetical social network indicated that they would share the story, implying that the majority of group members endorsed the story. See Figure 7 for an example of what two of these social network diagrams looked like. Participants then indicated how accurate they perceived the story to be on a 7-point scale ranging from "Not at all accurate" to "Very accurate," and indicated whether they would consider sharing this story online (with the possible responses "No," "Maybe," and "Yes," where "Yes" or "Maybe" was coded as 1 to form the variable "shared").

After this, participants indicated how accurate each of the five social connections in their social network were as judges, given their decision to share or not share the fake news, on a scale from 1 to 7. We were interested in examining whether social connections who conformed with other group members by sharing the story (e.g., Person E in the left-hand side of Figure 7) were perceived as more accurate than social connections that deviated by not sharing the story (e.g., Person B in the right-hand side of Figure 7).

For each news story, a participant was randomly assigned to either rate a conforming or a deviant social connection. Following this, participants responded to a two item scale measuring the extent to which they wanted to interact with this social connection in the future using a 7-point scale to rate agreement with the statements: "I want Person X to be in my social network" and "I want to follow the stories shared by Person X." We averaged the two scores to create a composite measure reflecting the participant's future desire to socially interact with this social connection ($\alpha = .91$). We repeated the same process for each of the six stories, after which participants provided demographic information (age, gender, education, and race) and political ideology with the same measures as in previous experiments.

Figure 7
Social Network Diagrams Used in Experimental Manipulation



Note. See the online article for the color version of this figure.

Results

We predicted a positive interaction such that when a social connection conformed to the group by sharing the fake news *and* the participant indicated they would share the story, the participant's desire for social interaction with the connection would be higher (lower social costs). Since participants were exposed to a battery of stories, some of which were not aligned with their political ideology, this interaction is analogous to the main effect of the social connection choosing whether or not to share in Studies 1–4, where the stories being considered had been chosen by respondents, so were presumably always endorsed.

Social connections who conformed to the group by sharing the story that the participant also shared were rated much more positively (in terms of accuracy and desire to socially interact with them) than the group members in the other three cells (see Figure S3 in the online supplemental materials for further details). As per our pre-registration, we analyzed the data with linear regression estimated with GEEs (Generalized Estimating Equations), finding a significant interaction effect between the social connection conforming and the participant choosing to share the falsehood in predicting both ratings of the social connection's accuracy ($b = 2.41$, $p < .001$) and participants' desire to socially interact with that social connection ($b = 1.92$, $p < .001$), which was also true without the inclusion of controls for age, gender, education, and political ideology (see the Table S11 in the online supplemental materials for further details). Overall, our hypothesis was supported: when the content was ideologically aligned, there were social costs for not sharing a story relative to sharing it.

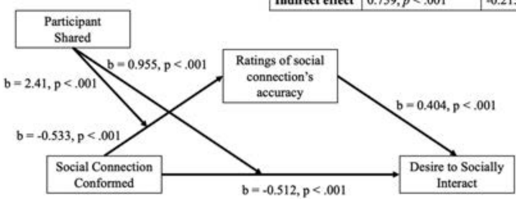
We also estimated a moderated mediation model examining if the interactive effect of the social connection's choice on whether to conform and the participant's choice on whether to share the story in predicting desire to socially interact was mediated by the subjective perception of the connection's accuracy as a judge (see Figure 8, Table S15 in the online supplemental materials).[10] The resulting analysis revealed that when considering a story that the participant would share, there was both a direct benefit of the friend conforming on desire to socially interact ($b = 0.444$, $p < .001$) and an indirect benefit via accuracy ($b = 0.759$, $p < .001$). Put differently, for those who did not conform, there was both a tribalistic (direct) reaction (37% of the total effect) and a rational evaluation of the group member's utility as an information source (i.e., the indirect effect via accuracy; 63% of the total effect).[11] Note that the tribalistic reaction could merely represent the judging group member aiming to impose penalties to signal loyalty to the group—rather than true devotion to the tribe (Willer et al., 2009).

[10] This was an additional moderated mediation model that was not pre-registered. Refer to Tables S12 and S14 for full estimates of the pre-registered models. In the pre-registered model, we did not indicate an intention to moderate the direct effect of conformity on desire to socially interact with whether the participant indicated they would share the story.

[11] When considering the penalty for a group member who conformed by sharing a story that the participant did *not* personally endorse, the direct effect was 70% of the total effect, and the indirect via accuracy 30%. This could suggest a more tribal punishment.

Figure 8
Moderated Mediation Model From Study 6a

	Participant shared	Participant did not share
Direct effect	0.444, $p < .001$	-0.512, $p < .001$
Indirect effect	0.759, $p < .001$	-0.215, $p < .001$



Discussion

Study 6a demonstrated that when a group exhibited consensus on sharing a story, respondents favorably evaluated group members who conformed with the group to share the story, provided they also personally would share the story. Both of these conditions were important: the judging group member had to personally endorse the story to impose costs for not sharing. This yields the insight that in homogenous groups where group members are likely to privately endorse the content being shared, deviant group members will face higher social costs. These higher social costs may further motivate group members to conform, thus increasing the spread of such falsehoods.

Study 6a complements the results of Studies 3–4 by showing that social costs for not sharing emerged in an experimental context with a modified paradigm that included reference to a wider group behavior, rather than a single social connection. Further analyses examining the relationship between deviance, ratings of group members' accuracy, and desire to interact with those group members revealed that people evaluated deviant group members both as less useful—in terms of their accuracy as judges of content—and separately, as less desirable to interact with. This decomposition suggests there are both rational and tribalistic elements to the observed social costs, adding further nuance to our understanding of the effect.

Despite these results, Study 6a suffered from two limitations: (a) participants were asked to imagine that their group chose to share different stories, some of which were misaligned with their own politics, which may not be realistic and (b) the hypothetical groups always chose to share the focal stories, which precludes gaining insight about the case when the group does not share a story. Regarding the first point, this was a choice made to preserve random assignment thereby allowing us further confidence in the causal role of deviance on social costs. In Studies 3 and 4, we allowed participants to choose stories (including the possibility of choosing none at all in Study 4), and found the same results. The fact that our conclusions are the same across approaches increases our confidence in the effect's robustness. We address the second limitation in the next study.

Study 6b

Study 6b examined the consequences for lone deviants who acted contrary to the group's behavior: to do so, we manipulated the lone deviant's behavior, varying whether they defied the group to not share supported content, or to share unsupported content. In short, we were interested in finding out if respondents would penalize group members who deviated from the group in ways aligned with the respondent's own beliefs. This study design further allowed us to look at the effects of manipulating the group's behavior on respondents' own sharing decisions.

Method

Participants

We recruited 493 participants from the online research participant platform Prolific. Participants were paid $1.51 for taking part in the survey. We removed six non US participants by manually checking their IP addresses.[12] The final sample contained 48.7% women, 50.3% men, and 1.0% non-binary people, with an average age of 34.3 years. Our final sample contained 143 Republicans and 344 Democrats. We pre-registered our sample size, hypothesis, and analysis (https://osf.io/k5a9q/).

Procedure

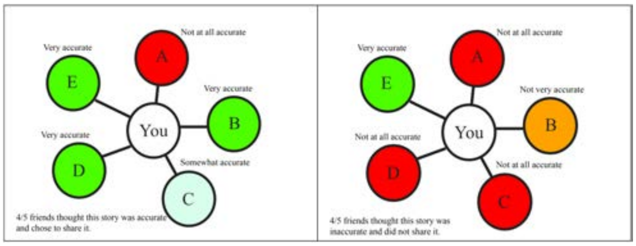
The study procedure was similar to Study 6a. Participants saw six different fake news stories with associated social network diagrams and initially indicated their perceptions of the accuracy of the fake stories and their likelihood of sharing them. However, there was one key difference compared to Study 6a. In Study 6a, the social network always endorsed the story and the manipulation was whether the participant rated a deviant or a conforming network member, whereas in Study 6b, the participants always rated a deviant member, and the manipulation was whether the lone deviant did not share the fake news when the network did, or shared it when the network did not. To be more specific, in one condition, four-fifths of the members of the network indicated that they would share the fake news. In the other condition, only one-fifth of the members of the network indicated that they would share the fake news. For this initial stage, the manipulation meant that participants either rated their own preferences for fake news whilst seeing a social network diagram that indicated their network generally shared the story, or one that showed their network generally did not share the story.

Participants then rated the accuracy of each social connection in the network as a judge, before answering questions about a specific social connection. In the four-fifths members' shared condition, the participant responded to questions regarding the lone deviant who did not share the fake news. Conversely, in the other condition, the participant rated the lone deviant who did share the fake news. See Figure 9 for two example networks—in the left-hand side panel the participant would be asked to rate Person A, and in the right hand side panel they would be asked to rate Person E.

For each of the 6 fake news stories a participant faced, they were randomly assigned to these conditions: either the "lone deviant did not share" condition (i.e., the left panel of Figure 9), or the "lone deviant shared" condition (the right panel) for each story, subject to the constraint that they faced three fake news stories in each condition. There were six different unique network diagrams in total where the lone deviant was sometimes A, B, C, and so on. After stating their own preference for sharing the fake news and the accuracy of other social connections as judges of the fake news, participants rated their

[12] Our bot identification methodology identified nine potential bots in this sample. We present our pre-registered analyses both with ($N = 487$) and without ($N = 478$) the inclusion of these potential bots, see the SI for further details.

Figure 9
Social Network Diagrams Used in Experimental Manipulation



Note. See the online article for the color version of this figure.

desire to socially interact with the lone deviant using the same two items from Study 6a ($\alpha = .89$). As an exploratory analysis, we also included two items to examine whether participants reported that the lone deviant should conform to the group (e.g., "Person A should act as a team player when it comes to the sharing of news stories"; $\alpha = .84$). After completing this process for each of the six fake stories, participants reported their demographic information as in earlier studies (age, gender, race, education, political ideology), but we measured partisan identification using a forced response between the Democrat and Republican party.[13]

Results

We examined the effects of our manipulation on the perceived accuracy of and their desire to socially interact with the deviant group member. Figure 10 plots the average ratings on these two variables. We observed a clear interaction pattern for both of our main variables. Lone deviants who either did not share a story that the participant did share—or did share a story that the participant did not share—were both evaluated negatively.[14] Essentially, there was a social cost for not sharing endorsed content (difference between blue bars) or sharing content that was not endorsed (difference between red bars).

Using the same regression procedure as Study 6a, we corroborated this observation, observing a significant negative interaction between the lone deviant not sharing and the participant sharing on their desire to socially interact with the group member—there was a significant interaction both without control variables ($b = -2.08$, $p < .001$) and with their inclusion ($b = -2.07$, $p < .001$), see Table S16 in the online supplemental materials for further details. This pattern was mirrored for perceptions of the social connection's accuracy (without control variables, $b = -2.64$, $p < .001$, with controls, $b = -2.64$, $p < .001$). In other words, when the participant endorsed the fake story, they wanted to interact less with people who did not, and saw them as less accurate than those who defied the group to share the story. Additional analysis—the same process of comparing the size of direct and indirect effects that we used in Study 6a—revealed that these social costs were more group-based in nature, as opposed to responses to a group member's reduced utility (when a participant endorsed a story, 51% of the social cost was associated with the direct effect compared to 37% in Study 6a; refer to the online supplemental materials for additional details).

While this was not the primary focus of Study 6b, the design we implemented also facilitated analyzing the effect of the group's behavior on the participants' decisions regarding whether to share stories. Participants saw either a group where the majority of the members shared a story (the left side of Figure 9) or a group where the majority of the members did not share (the right side of Figure 9), and were then asked whether they would like to share that story. In the case where the group chose to share, the participants would likely face higher social costs from failing to comply with the group to share the story. Thus, our conditions served as an experimental manipulation of the social costs that people faced. Indeed, we observed that network sharing had a positive effect on people's likelihood of sharing ($b = 0.390$, $p < .001$, Table S20 in the online supplemental materials), thus corroborating the results of Study 5 (in the SI, we present these results in full). Overall, we found evidence that social costs contributed to people's decisions to share falsehoods, both observationally in Study 5, and with an experimental manipulation in Study 6b.

Discussion

Lone deviants faced just as severe social costs for failing to share content that was endorsed by the participant as they were for sharing content not endorsed by the participant. This emphasizes how strong the costs are for failing to conform—they resemble the costs for going against the group to share ideologically opposed content. Here, our results complement those of Altay et al. (2022): we also find that there are reputational costs associated with sharing fake news, but only when the content is not endorsed by the judge. When the content is privately endorsed, there are social returns to sharing it; and social costs for failure to do so. In this context, group members do not blindly impose penalties for deviance from the group's behavior, which offers some hope. However, the social impact of real-world groups that consist of family and friends are likely to provide a much stronger motivation to conform, which could attenuate the role of private beliefs. Finally, Study 6b provided further support for Study 5's results by replicating that social costs were a potent psychological motivator of people's sharing decisions, and should be considered along with partisanship and subjective accuracy beliefs as a key factor that contributes to the spread of misinformation.

General Discussion

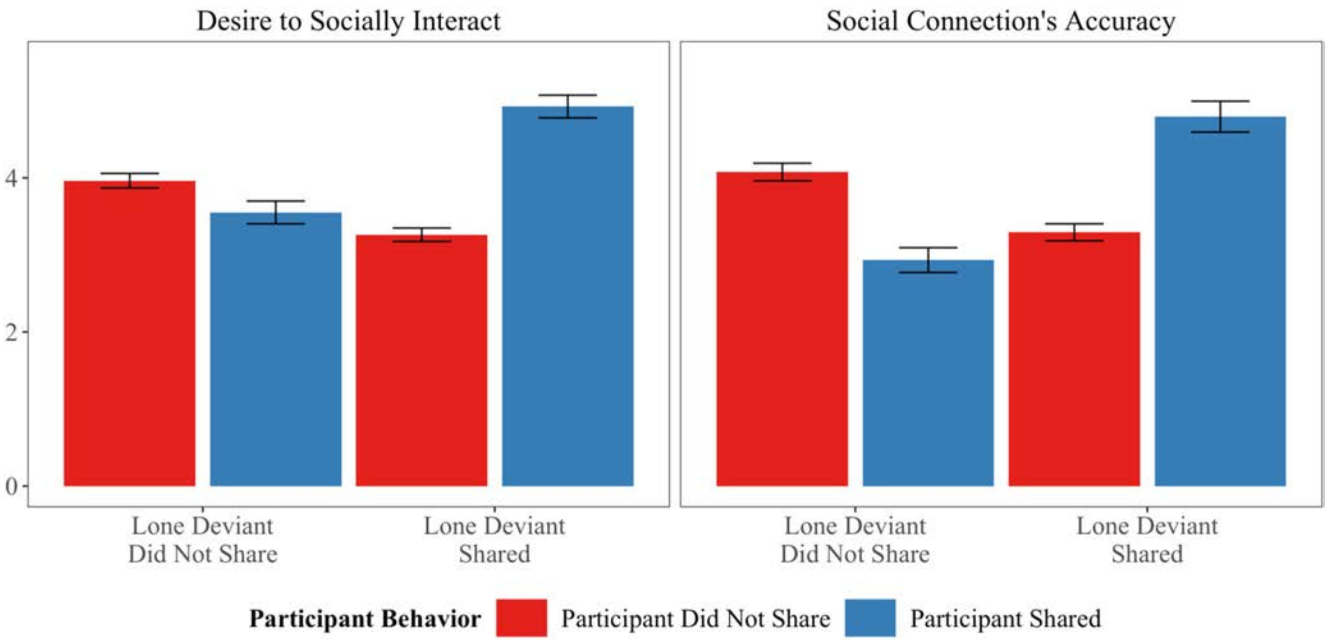
The spread of misinformation is a serious and escalating problem, which underscores the need to understand why people share it. In the present research, we investigated the impact of people's social contexts on their decisions regarding whether to share fake news. We found significant social costs associated with failing to share fake news that was endorsed by other group members, demonstrating the role of group conformity as a key motivator to spread falsehoods.

[13] We moved away from this binary forced choice in studies conducted temporally later (e.g., Study 6a) after receiving feedback to allow for the possibility of political independents.

[14] In our pre-registration, we stated our intention to use the political concordance of the stories as our independent variable (i.e., whether the participant indicated they preferred the Republican party and the story was Republican-leaning), but later determined that participants' choices regarding whether they would share each specific story was a more proximal measure of their endorsement. We repeat our analyses with the originally planned measure of endorsement in the supplementary information. Our conclusions are unchanged.

Figure 10

Average Ratings on Social Connection Accuracy and Desire to Socially Interact with Social Connections for Shared and Not Shared Stories and Lone Deviants Who Shared and Did Not Share Stories



Note. Error bars signify 95% confidence intervals. In each panel, the bars from left to right detail "Participant Did Not Share" for "Lone Deviant Did Not Share", "Participant Shared" for "Lone Deviant Did Not Share", "Participant Did Not Share" for "Lone Deviant Shared", and "Participant Shared" for "Lone Deviant Shared". See the online article for the color version of this figure.

Using Twitter users' real behavior, Study 1 found that not sharing a fake story that a group member shared was associated with a significantly greater decrease in social interaction over time. Study 2 built on this, showing that the relationship between sharing content in common and social interaction was significantly stronger within the fake news ecosystem relative to a more representative sample of Twitter users. We found causal support for this relationship in Studies 3 and 4—participants indicated a reduced desire to socially interact with social connections who failed to share the same falsehoods as them. Study 4 further built on the results of Study 2, revealing that these social costs were stronger specifically for not sharing fake news. Study 5 demonstrated that people both recognized these social costs and that they were associated with the likelihood of sharing falsities, suggesting social costs as a psychological motivator of fake news dissemination. Finally, Studies 6a and 6b revealed that both deviances from the group (i.e., failure to share) and the judging group member's own agreement with a story are key factors in determining the severity of social costs. Taken together, this research documents the role of group pressure and conformity in the sharing of fake news.

Theoretical Contributions

Our research makes several important contributions. First, our work underscores the role of conformity as a key psychological motivator of sharing fake news. People's aversion to facing social costs may lead them to share falsehoods. This highlights the importance of considering social affiliation needs as a predictor of spreading fake news. Existing research has mostly documented the role of social and group identities in cases where individuals who identify with a certain group are motivated to share fake as a means to enhance their group's image or derogate members of their out-group (Van Bavel et al., 2021). With such identities, the motivation to help the group is triggered at the individual level. In contrast, we show that the impetus for conformity is triggered by the actions of other group members, who impose social costs to elicit acquiescence to sharing falsehoods. As a result, our work not only contributes to the literature by examining the relatively underexplored role of group-level factors in the sharing of fake news (Scheufele & Krause, 2019), but also offers a perspective that complements existing research on the role of social identities.

Second, we found that social costs are a mechanism by which conformity leads to sharing fake news. People fearing social costs are more likely to share fake news. Importantly, this fear is warranted: Social connections who did not share the same fake news as others experienced reduced social interaction on Twitter. Likewise, participants indicated a tendency to sideline deviant group members for future social interactions in the experimental studies. By documenting both the role of social costs in determining sharing decisions and their real presence on Twitter, our work advances the literature by identifying social costs as a key psychological driver of the spread of misinformation. We also established that social costs are both a unique predictor above and beyond existing explanations surrounding partisan identity and accuracy and a pathway by which partisan identity can influence sharing decisions.

Third, our work sheds light on how online groups on social media platforms can become echo chambers that reinforce increasingly polarized views. Recent work has demonstrated that smaller online groups are a more potent driver of fake news than larger ones (Vosoughi et al., 2018). Our work offers a potential explanation for this empirical finding: the role of conformity pressure and social costs within small online clusters. We found evidence of social costs both for failure to share endorsed content (Studies 1–4) and for sharing unendorsed content (Studies 6a and 6b), corroborating the evidence of a recent study (Altay et al., 2022). This suggests that these social costs directly impede the diversity of perspectives in groups. Overall, we argue that the social and reputational effects of sharing fake news depend on the norms in a group and the group's epistemological reality. It is likely that groups will punish members for sharing content that contravenes their existing belief system, and reward those who maintain it. Due to this, group members' sharing may go beyond expressive responding (Schaffner & Luks, 2018) to reflect responding that is designed to zealously maintain one's status in the group and the group's constructed reality, regardless of the truth (Willer et al., 2009).

Additionally, we found that the social costs for failing to support group members by sharing the same content were not distributed equally. Social costs were specifically higher for not sharing fake news compared to other news. This was not only observed in the field studies on Twitter (Studies 1 and 2) but also in an experiment (Study 4). This further suggests the importance of considering social costs when trying to understand fake news as a phenomenon. The social costs were also higher in groups closely associated with conservative political ideology. For instance, social costs were a pathway by which Republican partisan identity elevated the likelihood of sharing falsehoods (Study 5), and political conservatives imposed harsher social costs on group members (Study 1). These findings increase our understanding of why political conservatives may be more prone to sharing fake news (e.g., Lawson & Kakkar, 2021).

Finally, our work suggests that interventions that target individuals' social affiliation motives might be an effective approach to reduce the spread of misinformation. There is a wealth of evidence that existing methods focused on "pre-bunking"—that is, informing people of the ways in which they may be misinformed as a means of inoculating them against susceptibility to future misinformation—are an effective intervention for reducing the spread of fake news (e.g., Roozenbeek et al., 2022). Another research tradition focuses on the role of accuracy primes, showing that drawing people's attention to the accuracy of news may reduce the spread of falsehoods (Pennycook et al., 2021; Pennycook & Rand, 2022), though there is some debate over their efficacy for Republicans (Rathje et al., 2022). Given the evidence documented here that social costs affect fake news sharing via a unique pathway, this could point to a fruitful opportunity for a third pathway to reducing susceptibility to misinformation: designing socially affirming interventions that attenuate the perceived risk of social costs. Future work might examine how the fear of social costs can be attenuated and the implications of such interventions for regulation surrounding social media.

Limitations

One possible limitation of our research is the lack of causal evidence in the field studies. We conclude from the overall evidence that it is likely that not sharing falsehoods leads to social costs, but this requires triangulating across externally valid correlational evidence and internally valid experimental evidence. It could be the case that the causality we observed in our randomly assigned experiments does not apply in the applied context. Future research could seek to identify a discontinuity or appropriate instrument to recover a causal estimate of the effect of failing to share falsehoods. Relatedly, we relied on correlational mediation analysis to test our proposed mechanisms. This method has well-documented shortcomings (Bullock et al., 2010; D. P. Green et al., 2010; Imai et al., 2011; Rohrer, 2018) and so further evidence is required to conclusively validate some of our results regarding the mechanisms.

Further, research on misinformation is highly dependent on the fake news stimuli used to investigate the question of interest. These stimuli (i.e., stories) are more heterogeneous than other decision-making research topics (e.g., decision-making under risk), which can lead to them having a disproportionate impact on the conclusions one can draw from a study. To combat this, we ran two Twitter field studies, one which focused on people sharing content from 974 fake and hyperpartisan news websites (Study 1), and one which focused on different kinds of people sharing any type of content in common (Study 2). However, the conclusions drawn from our experiments may suffer from this stimuli specificity problem that is pervasive in misinformation research. We maximized the variety of our approaches, using archival studies, different samples, and different stimuli, but the conclusions are likely to be less general than those that depend on more homogeneous stimuli.

Relatedly, a particular limitation of research that uses fake news stimuli is that these are highly reflective of the current cultural context, and the cultural context is always in flux. Given this fact, changes in the actions of political elites have profound impacts on how misinformation is produced and interpreted (Van Bavel et al., 2021). The most salient example of this is the banning of Donald Trump from Twitter, which had knock-on effects for the entire social media ecosystem (Dwoskin & Timberg, 2021); Elon Musk's 2022 purchase of Twitter may have similar consequences. These factors have implications for how we consider replication and reproducibility in the case of misinformation. For example, research conducted prior to the 2020 Presidential election may not replicate in the period after the election if it relies on stimuli that relied on Donald Trump being president. This does not necessarily compromise the generality of the phenomenon, but does require using equivalent rather than identical stimuli in replication attempts, which introduces researcher degrees of freedom.

Future Directions

One fruitful avenue for future research is further understanding the role of political ideology and partisan identity in determining social costs. In our data, we observed conflicting results regarding the role of ideology: conservative political ideology was associated with greater social costs in Study 1, but this relationship did not hold up in our experiments. One possible reason for this inconsistency could be that conservative Twitter users tend to end up in more ideologically skewed networks (Chen et al., 2021). It could also be a result of differences in the methodology (including the social desirability concerns and smaller samples of surveys) or population—people sharing fake news in the wild are likely to be psychologically atypical. Regardless, it warrants further investigation.

A related phenomenon that requires further study is considering who is the appropriate population in which to study misinformation sharing. The gold standard for psychological research is generally to look at Nationally representative samples, but fake news sharing is concentrated in a small proportion of the population, and general samples contain few of these people who actually share fake news. Should we strive to study the behavior of those who do share fake news, or compare them to a broader set to establish *why* these people are the ones sharing falsehoods? One takeaway from the present research is the importance of pursuing field and laboratory experiments in conjunction, to study both commonalities and differences that may require further explanation.

Another future direction concerns distinguishing between the effects of news content and the differences in the kinds of people who share different content. For example, we found that social costs were stronger for not sharing news that was specifically fake. The people who choose to share fake news are surely different to those who choose to share real news, which begs the question: Are the observed differences attributable to the nature of the news, or selection effects in who would share such news? Further research is required to understand the relative contributions of the nature of the *content* and the nature of the *people* who share such content to decisions regarding fake news and social penalties.

Another key question is: exactly what features of the fake and real news determines people's behavior toward them? Many factors have been held up as explaining the spread of content online, including novelty and emotionality (Milkman & Berger, 2014), outgroup animosity (Rathje et al., 2021), and political partisanship (Osmundsen et al., 2021). This invites the possibility that people could behave differently toward fake news for reasons other than its veracity. In fact, given that veracity is often not known—due to a turbulent information environment and the bias of news sources—it is not only possible, but likely that some other axis on which fake and real news differs explains the observed differences. It is possible that our effects—where people receive social penalties for failing to share falsehoods—would replicate for other types of content that people care about. Indeed, we do see this in Study 2, but still find that the effect is stronger for specifically fake news.

Finally, given the divergence between people's accuracy beliefs and sharing intentions (Pennycook et al., 2021), it is important to disentangle whether social penalties only foster compliance, or whether this compliance is ultimately internalized and group members come to view misinformation as accurate. This is important in determining the best strategy to combat people feeling pressured to share fake news.

Conclusion

Using ecologically valid field data and five experiments, this research found widespread evidence that conformity pressure and social costs are a key psychological driver of sharing fake news. Social costs both independently contributed to sharing decisions and explained the relationship between partisanship and sharing. In drawing attention to the social costs suffered by deviant group members, our work reveals a contributing factor to how online tribes are formed and maintained, and how misinformation can become a unifying group culture.

References

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

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, *31*(2), 211–236. https://doi.org/10.1257/jep.31.2.211

Allen, J., Howland, B., Mobius, M., Rothschild, D., & Watts, D. J. (2020). Evaluating the fake news problem at the scale of the information ecosystem. *Science Advances*, *6*(14), Article eaay3539. https://doi.org/10.1126/sciadv.aay3539

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

Altemeyer, B. (1981). *Right-wing authoritarianism*. University of Manitoba Press.

Asch, S. E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological Monographs: General and Applied*, *70*(9), 1–70. https://doi.org/10.1037/h0093718

Bago, B., Rand, D. G., & Pennycook, G. (2020). Fake news, fast and slow: Deliberation reduces belief in false (but not true) news headlines. *Journal of Experimental Psychology: General*, *149*(8), 1608–1613. https://doi.org/10.1037/xge0000729

Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, *348*(6239), 1130–1132. https://doi.org/10.1126/science.aaa1160

Bandura, A. (2000). Exercise of human agency through collective efficacy. *Current Directions in Psychological Science*, *9*(3), 75–78. https://doi.org/10.1111/1467-8721.00064

Barberá, P. (2015). Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data. *Political Analysis*, *23*(1), 76–91. https://doi.org/10.1093/pan/mpu011

Batailler, C., Brannon, S. M., Teas, P. E., & Gawronski, B. (2022). A signal detection approach to understanding the identification of fake news. *Perspectives on Psychological Science*, *17*(1), 78–98. https://doi.org/10.1177/1745691620986135

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, *117*(3), 497–529. https://doi.org/10.1037/0033-2909.117.3.497

Bettenhausen, K. L., & Murnighan, J. K. (1985). The emergence of norms in competitive decision-making groups. *Administrative Science Quarterly*, *30*(3), 350–372. https://doi.org/10.2307/2392667

Bettenhausen, K. L., & Murnighan, J. K. (1991). The development of an intragroup norm and the effects of interpersonal and structural challenges. *Administrative Science Quarterly*, *36*(1), 20–35. https://doi.org/10.2307/2393428

Bullock, J. G., Green, D. P., & Ha, S. E. (2010). Yes, but what's the mechanism? (don't expect an easy answer). *Journal of Personality and Social Psychology*, *98*(4), 550–558. https://doi.org/10.1037/a0018933

Caporael, L. R. (1997). The evolution of truly social cognition: The core configurations model. *Personality and Social Psychology Review*, *1*(4), 276–298. https://doi.org/10.1207/s15327957pspr0104_1

Centola, D., Willer, R., & Macy, M. (2005). The emperor's dilemma: A computational model of self-enforcing norms. *American Journal of Sociology*, *110*(4), 1009–1040. https://doi.org/10.1086/427321

Chen, W., Pacheco, D., Yang, K.-C., & Menczer, F. (2021). Neutral bots probe political bias on social media. *Nature Communications*, *12*(1), Article 5580. https://doi.org/10.1038/s41467-021-25738-6

Cialdini, R. B., & Trost, M. R. (1998). Social influence: Social norms, conformity and compliance. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (pp. 151–192). McGraw-Hill.

Cinelli, M., De Francisci Morales, G., Galeazzi, A., Quattrociocchi, W., & Starnini, M. (2021). The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, *118*(9), Article e2023301118. https://doi.org/10.1073/pnas.2023301118

Correll, J., & Park, B. (2005). A model of the Ingroup as a social resource. *Personality and Social Psychology Review*, *9*(4), 341–359. https://doi.org/10.1207/s15327957pspr0904_4

Darwin, C. (1896). *The descent of man and selection in relation to sex*. D. Appleton.

Douglas, K. M., Sutton, R. M., & Cichocka, A. (2017). The psychology of conspiracy theories. *Current Directions in Psychological Science*, *26*(6), 538–542. https://doi.org/10.1177/0963721417718261

Dwoskin, E., & Timberg, C. (2021, January 16). Misinformation dropped dramatically the week after Twitter banned Trump and some allies. *Washington Post*. https://www.washingtonpost.com/technology/2021/01/16/misinformation-trump-twitter/

Ecker, U. K. H., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., Kendeou, P., Vraga, E. K., & Amazeen, M. A. (2022). The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology*, *1*(1), Article 1. https://doi.org/10.1038/s44159-021-00006-y

Egebark, J., & Ekström, M. (2011). *Like what you like or like what others like? Conformity and peer effects on Facebook* [SSRN Scholarly Paper ID 1948802]. Social Science Research Network. https://doi.org/10.2139/ssrn.1948802

Epstein, Z. (2018). *Fake news list* [Jupyter Notebook]. https://github.com/zivepstein/fake-news-list

Erikson, K. T. (1966). *Wayward puritans: A study in the sociology of deviance*. Wiley and Sons.

Fehr, E., & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior*, *25*(2), 63–87. https://doi.org/10.1016/S1090-5138(04)00005-4

Frenkel, S., Alba, D., & Zhong, R. (2020, March 8). Surge of virus misinformation stumps Facebook and Twitter. *The New York Times*, *19*. https://www.nytimes.com/2020/03/08/technology/coronavirus-misinformation-social-media.html

Friggeri, A., Adamic, L., Eckles, D., & Cheng, J. (2014). Rumor cascades. *Proceedings of the International AAAI Conference on Web and Social Media*, *8*(1), Article 1. https://doi.org/10.1609/icwsm.v8i1.14559

Green, D. P., Ha, S. E., & Bullock, J. G. (2010). Enough already about "black box" experiments: Studying mediation is more difficult than most scholars suppose. *The Annals of the American Academy of Political and Social Science*, *628*(1), 200–208. https://doi.org/10.1177/0002716209351526

Green, J., Hobbs, W., McCabe, S., & Lazer, D. (2022). Online engagement with 2020 election misinformation and turnout in the 2021 Georgia runoff election. *Proceedings of the National Academy of Sciences*, *119*(34), Article e2115900119. https://doi.org/10.1073/pnas.2115900119

Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., & Lazer, D. (2019). Fake news on Twitter during the 2016 U.S. presidential election. *Science*, *363*(6425), 374–378. https://doi.org/10.1126/science.aau2706

Guess, A. M. (2021). (Almost) everything in moderation: New evidence on Americans' online media diets. *American Journal of Political Science*, *65*(4), 1007–1022. https://doi.org/10.1111/ajps.12589

Guess, A. M., Nyhan, B., & Reifler, J. (2020). Exposure to untrustworthy websites in the 2016 US election. *Nature Human Behaviour*, *4*(5), 472–480. https://doi.org/10.1038/s41562-020-0833-x

Hogg, M. A. (2001). A social identity theory of leadership. *Personality and Social Psychology Review*, *5*(3), 184–200. https://doi.org/10.1207/S15327957PSPR0503_1

Hollander, E. P. (1958). Conformity, status, and idiosyncrasy credit. *Psychological Review*, *65*(2), 117–127. https://doi.org/10.1037/h0042501

Huckfeldt, R., Mendez, J. M., & Osborn, T. (2004). Disagreement, ambivalence, and engagement: The political consequences of heterogeneous networks. *Political Psychology*, *25*(1), 65–95. https://doi.org/10.1111/j.1467-9221.2004.00357.x

Imai, K., Keele, L., Tingley, D., & Yamamoto, T. (2011). Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies. *The American Political Science Review*, *105*(4), 765–789. https://doi.org/10.2307/23275352

Jervis, R., Ramirez, M., & Ruiz-Goiriena, R. (2021, January 12). Evangelical Christians back Trump remarks on capitol riot and Biden. *USA Today*. https://www.usatoday.com/story/news/nation/2021/01/12/evangelicals-donald-trump-capitol-riot-voter-fraud/6644005002/

John, M. (2021, January 13). REUTERS NEXT public trust crumbles under COVID-19, fake news-survey. *Reuters*. https://www.reuters.com/business/media-telecom/reuters-next-public-trust-crumbles-under-covid-19-fake-news-survey-2021-01-13/

Jost, J. T., van der Linden, S., Panagopoulos, C., & Hardin, C. D. (2018). Ideological asymmetries in conformity, desire for shared reality, and the spread of misinformation. *Current Opinion in Psychology*, *23*, 77–83. https://doi.org/10.1016/j.copsyc.2018.01.003

Juul, J. L., & Ugander, J. (2021). Comparing information diffusion mechanisms by matching on cascade size. *Proceedings of the National Academy of Sciences*, *118*(46), Article e2100786118. https://doi.org/10.1073/pnas.2100786118

Kahan, D. M. (2017). *Misconceptions, misinformation, and the logic of identity-protective cognition* [SSRN Scholarly Paper ID 2973067]. Social Science Research Network. https://doi.org/10.2139/ssrn.2973067

Kang, C., & Goldman, A. (2016, December 5). In Washington pizzeria attack, fake news brought real guns. *The New York Times*. https://www.nytimes.com/2016/12/05/business/media/comet-ping-pong-pizza-shooting-fake-news-consequences.html

Kivran-Swaine, F., Govindan, P., & Naaman, M. (2011). *The impact of network structure on breaking ties in online social networks: Unfollowing on Twitter* 4.

Kuran, T. (1995). The inevitability of future revolutionary surprises. *American Journal of Sociology*, *100*(6), 1528–1551. https://www.journals.uchicago.edu/doi/abs/10.1086/230671

Lawson, M. A., & Kakkar, H. (2020, June 23). *Tribalism and tribulations: The social costs of not sharing fake news*. OSF. https://osf.io/qbkzu/

Lawson, M. A., & Kakkar, H. (2021). Of pandemics, politics, and personality: The role of conscientiousness and political ideology in the sharing of fake news. *Journal of Experimental Psychology: General*, *151*(5), 1154–1177. https://doi.org/10.1037/xge0001120

Lerman, K., Yan, X., Wu, X.-Z., & Amblard, F. (2016). The "majority illusion" in social networks. *PLoS ONE*, *11*(2), Article e0147617. https://doi.org/10.1371/journal.pone.0147617

Lewandowsky, S. (2022). Fake news and participatory propaganda. In R. F. Pohl (Ed.), *Cognitive illusions* (3rd ed., pp. 324–340). Routledge.

Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest*, *13*(3), 106–131. https://doi.org/10.1177/1529100612451018

Lewandowsky, S., Jetter, M., & Ecker, U. K. H. (2020). Using the president's Tweets to understand political diversion in the age of social media. *Nature Communications*, *11*(1), Article 1. https://doi.org/10.1038/s41467-020-19644-6

Liefbroer, A. C., & Billari, F. C. (2010). Bringing norms back in: A theoretical and empirical discussion of their importance for understanding demographic behaviour. *Population, Space and Place*, *16*(4), 287–305. https://doi.org/10.1002/psp.552

Manganelli Rattazzi, A. M., Bobbio, A., & Canova, L. (2007). A short version of the Right-Wing Authoritarianism (RWA) Scale. *Personality and Individual Differences*, *43*(5), 1223–1234. https://doi.org/10.1016/j.paid.2007.03.013

Marini, M. M. (1984). Age and sequencing norms in the transition to adulthood*. *Social Forces*, *63*(1), 229–244. https://doi.org/10.1093/sf/63.1.229

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, *27*(1), 415–444. https://doi.org/10.1146/annurev.soc.27.1.415

Milkman, K. L., & Berger, J. (2014). The science of sharing and the sharing of science. *Proceedings of the National Academy of Sciences*, *111*(Suppl 4), 13642–13649. https://doi.org/10.1073/pnas.1317511111

OECD. (2020, July 3). *Transparency, communication and trust: The role of public communication in responding to the wave of disinformation about the new coronavirus*. https://www.oecd.org/coronavirus/policy-responses/transparency-communication-and-trust-the-role-of-public-communication-in-responding-to-the-wave-of-disinformation-about-the-new-coronavirus-bef7ad6e/

O'Gorman, H. J., & Garry, S. L. (1976). Pluralistic ignorance—A replication and extension. *Public Opinion Quarterly*, *40*(4), 449. https://doi.org/10.1086/268331

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

Pennycook, G., Cannon, T. D., & Rand, D. G. (2018). Prior exposure increases perceived accuracy of fake news. *Journal of Experimental Psychology: General*, *147*(12), 1865–1880. https://doi.org/10.1037/xge0000465

Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. (2021). Shifting attention to accuracy can reduce misinformation online. *Nature*, *592*, 590–595. https://doi.org/10.31234/osf.io/3n9u8

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

Pennycook, G., & Rand, D. G. (2019a). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, *188*, 39–50. https://doi.org/10.1016/j.cognition.2018.06.011

Pennycook, G., & Rand, D. G. (2019b). Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. *Journal of Personality*, *88*(2), 185–200. https://doi.org/10.1111/jopy.12476

Pennycook, G., & Rand, D. G. (2022). Accuracy prompts are a replicable and generalizable approach for reducing the spread of misinformation. *Nature Communications*, *13*(1), Article 1. https://doi.org/10.1038/s41467-022-30073-5

Petersen, M. B., Osmundsen, M., & Arceneaux, K. (in press). *A "Need for Chaos" and the sharing of hostile political rumors in advanced democracies. American Political Science Review*. https://doi.org/10.31234/osf.io/6m4ts

Rathje, S., Roozenbeek, J., Traberg, C., Van Bavel, J., & van der Linden, S. (2022). Letter to the editors of psychological science: Meta-analysis reveals that accuracy nudges have little to no effect for U.S. conservatives: Regarding Pennycook et al. (2020). *Psychological Science*. Advance online publication. https://doi.org/10.25384/SAGE.12594110.v2

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*, *118*(26), Article e2024292118. https://doi.org/10.1073/pnas.2024292118

Ren, Z. (Bella), Dimant, E., & Schweitzer, M. E. (2021). *Social motives for sharing conspiracy theories* [SSRN Scholarly Paper]. https://doi.org/10.2139/ssrn.3919364

Ridgeway, C. L., & Berger, J. (1986). Expectations, legitimation, and dominance behavior in task groups. *American Sociological Review*, *51*(5), 603–617. https://doi.org/10.2307/2095487

Rohrer, J. M. (2018). Thinking clearly about correlations and causation: Graphical causal models for observational data. *Advances in Methods and Practices in Psychological Science*, *1*(1), 27–42. https://doi.org/10.1177/2515245917745629

Roozenbeek, J., van der Linden, S., Goldberg, B., Rathje, S., & Lewandowsky, S. (2022). Psychological inoculation improves resilience against misinformation on social media. *Science Advances*, *8*(34), Article eabo6254. https://doi.org/10.1126/sciadv.abo6254

Schaffner, B. F., & Luks, S. (2018). Misinformation or expressive responding? What an inauguration crowd can tell us about the source of political misinformation in surveys. *Public Opinion Quarterly*, *82*(1), 135–147. https://doi.org/10.1093/poq/nfx042

Scheufele, D. A., & Krause, N. M. (2019). Science audiences, misinformation, and fake news. *Proceedings of the National Academy of Sciences*, *116*(16), 7662–7669. https://doi.org/10.1073/pnas.1805871115

Settersten, R. A., & Hagestad, G. O. (1996). What's the latest? II. Cultural age deadlines for educational and work transitions. *The Gerontologist*, *36*(5), 602–613. https://doi.org/10.1093/geront/36.5.602

Sherif, M. (1936). *The psychology of social norms* (pp. xii, 210). Harper.

Tajfel, H., & Turner, J. C. (1985). The social identity theory of intergroup behavior. In S. Worchel & W. G. Austin (Eds.), *Psychology of intergroup relations* (2nd ed. pp. 7–24). Nelson-Hall.

Van Bavel, J. J., Harris, E. A., Pärnamets, P., Rathje, S., Doell, K. C., & 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

Vandello, J., & Cohen, D. (2003). Male honor and female fidelity: Implicit cultural scripts that perpetuate domestic violence. *Journal of Personality and Social Psychology*, *84*(5), 997–1010. https://doi.org/10.1037/0022-3514.84.5.997

van der Linden, S. (2022). Misinformation: Susceptibility, spread, and interventions to immunize the public. *Nature Medicine*, *28*(3), 460–467. https://doi.org/10.1038/s41591-022-01713-6

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

Waruwu, B. K., Tandoc, E. C., Duffy, A., Kim, N., & Ling, R. (2021). Telling lies together? Sharing news as a form of social authentication. *New Media and Society*, *23*(9), 2516–2533. https://doi.org/10.1177/1461444820931017

Willer, R. (2005). *Overdoing gender: A test of the masculine overcompensation thesis*. Paper presented at the annual meeting of the American Sociological Association, Philadelphia.

Willer, R., Kuwabara, K., & Macy, M. W. (2009). The false enforcement of unpopular norms. *American Journal of Sociology*, *115*(2), 451–490. https://doi.org/10.1086/599250

Williams, K. D. (2007). Ostracism. *Annual Review of Psychology*, *58*(1), 425–452. https://doi.org/10.1146/annurev.psych.58.110405.085641

Wooldridge, J. M. (2019). *Introductory econometrics: A modern approach*. Cengage Learning.

Received September 8, 2022
Revision received January 5, 2023
Accepted January 8, 2023 ■