How might a dislike option affect how people evaluate and engage with online content?

Nicolas Fay[1]*, Bradley Walker[1], Toby Prike[2], Ullrich Ecker[1], Lucy Butler[3], Piers Howe[4], Mengbin Ye[5]

[1]School of Psychological Science, University of Western Australia; Perth, Australia

[2]School of Psychology, University of Adelaide; Adelaide, Australia

[3]Department of Psychology, Network Science Institute, Department of Political Science, Northeastern University; Boston, Massachusetts, United States

[4]School of Psychological Sciences, University of Melbourne; Melbourne, Australia

[5]School of Computer and Mathematical Sciences, University of Adelaide; Adelaide, Australia

Running Head: Dislike Option and Online Engagement

Keywords: Social Influence, Social Endorsement, Veracity, Misinformation, Information Sharing, Social Media, Engagement Metrics

*Corresponding author:
Nicolas Fay, School of Psychological Science, University of Western Australia
35 Stirling Highway, Crawley, WA 6009 Australia
Email: nicolas.fay@gmail.com; Tel: +61 (0)8 6488 2688; Fax: +61 (0)8 6488 1006

Word count (excluding title page, abstract, methods, figures captions & references): XXXX words

Word count (excluding abstract and references): 8205 words

Abstract

This research investigated how people respond to true and false news, and how the type of social endorsement associated with news posts influences belief and engagement intentions. Across three experiments (total *N* = 1084), participants viewed true and false news accompanied by varying levels of positive (likes) and negative (dislikes) endorsement. Participants were sensitive to post veracity, rating true news as more believable (large effect), and showing greater willingness to like true content, dislike false content and share true content (small effects). Positive social endorsement by others reliably increased belief, like and share intentions and decreased dislike intentions (small effects). Negative endorsement reduced belief (Experiment 1) and like intentions and increased dislike intentions (small effects), but had no detectable impact on share intentions. When negative endorsement was presented alongside high positive endorsement, this reduced belief and like intentions, but only when the negative endorsement level was high, signalling divided opinion. These findings demonstrate that: 1) veracity has a strong effect on beliefs, and a small effect on engagement and 2) social cues—positive and negative—have a small but significant role in shaping how people evaluate and interact with online news. Our findings suggest that incorporating a dislike option on social media platforms could improve content evaluation by signalling epistemic rejection, though safeguards may be needed to mitigate risks of misuse.

Introduction

The introduction of the 'Like' button in 2009 and the later addition of the 'Re-Share' function fundamentally changed social media by incentivising rapid, emotionally charged engagement at the expense of reasoned discourse (Haidt, 2022, 2025). Today, nearly all major social media platforms—from Facebook and Instagram to TikTok and X (formerly Twitter)—allow users to publicly endorse posts through likes, hearts, or equivalent signals. Rarely do they provide users with the option to publicly express disapproval, such as a 'Dislike' option (e.g., Reddit). This asymmetry in user feedback may contribute to the influence and spread of false information and other harmful content (Juul & Ugander, 2021; Vosoughi et al., 2018), which platform algorithms often amplify based on engagement metrics rather than accuracy or social value (Cinelli et al., 2021; McLoughlin & Brady, 2024). This paper investigates whether introducing a negative endorsement mechanism—allowing users to visibly express disapproval—affects how people evaluate and engage with online content, and whether such a mechanism may help curb the impact of harmful content by providing users and algorithms with more balanced signals of content quality.

Cultural evolutionary theory has identified a range of evolved cognitive biases that shape human social learning, broadly categorised as content-dependent and context-dependent biases (Boyd & Richerson, 1985; Kendal et al., 2018). Of particular relevance to this study are two well-established biases: a content bias favouring negative information (Baumeister et al., 2001; Rozin & Royzman, 2001) and a context bias favouring majority opinion (Asch, 1951; see also the bandwagon effect; Bindra et al., 2022). That is, people tend to believe, remember, and share negative information more readily than positive information (Bebbington et al., 2017; Fessler et al., 2014), and often conform to majority opinions, even when those opinions are objectively incorrect (Asch, 1955; Franzen & Mader, 2023), an effect also observed in non-human primates (Haun et al., 2012). These biases may reflect evolutionarily conserved, adaptive mechanisms that support threat detection and social cohesion, and their interaction may be especially potent in online environments. On social media, a negative endorsement metric showing the number of dislikes associated with a post could exploit both biases, with its inherent negativity attracting engagement and the number of dislikes signalling oppositional consensus and amplifying the influence of the negative endorsement. In this way, a negative endorsement mechanism may have a powerful influence on how people evaluate and engage with online content.

Reflecting the design of most social media platforms, researchers have primarily investigated how positive social endorsement—through likes and shares—influences belief formation and sharing behaviour. The general finding is that positive endorsement from many unknown others carries as much persuasive weight as endorsement from expert sources (Jucks & Thon, 2017); high levels of positive endorsement increase belief in and sharing of false information (Avram et al., 2020; Butler et al., 2022; Luo et al., 2022; Shin, 2022) and enhance the effectiveness of corrections to false information (Vlasceanu &

Coman, 2022). Related work suggests these effects are explained by the perceived consensus implied by visible positive endorsement (Lewandowsky et al., 2019; Traberg et al., 2024). By contrast, research on the impact of negative social endorsement (e.g., dislikes) is limited. To date, only one study has examined how relative endorsement affects belief in false information and its correction (Butler, Fay, et al., 2024). It found that posts accompanied by high relative positive endorsement (i.e., many likes and few dislikes) increased belief in false information compared to those accompanied by high relative negative endorsement (i.e., few likes and many dislikes), with minimal effect on correction effectiveness.

While these studies prioritise ecological validity by focusing on common features of social media platforms, it is important to note that negative endorsement options—though currently rare—would be simple to implement on social media platforms and could improve individual decision-making and promote a healthier information ecosystem. This possibility is tested across three experiments[1] that examine how people evaluate and engage with true and false news content, and how their responses are influenced by positive and negative social endorsement. In each experiment, we queried participants' real-world social media use, in part to validate our measure of online sharing intentions. We then examined participants' belief in and engagement intentions toward true and false information. Research shows that message veracity exerts a strong influence on persuasion, but a weak influence on sharing behaviour, which is primarily driven by social connection rather than by truth (Fay et al., 2025). This pattern aligns with the concept of *epistemic vigilance* (Sperber et al., 2010)—that people are equipped with a suite of cognitive mechanisms to evaluate the credibility of information and guard against being misinformed by others. We therefore predict that veracity will have a large effect on belief but a small effect on sharing intentions. While the effect of veracity on endorsement intentions is less clear, sharing and liking show a strong positive correlation (Tenenboim, 2022). We therefore predicted that people would be more inclined to like true content and dislike false content. Finally, in each experiment, we tested the extent to which others' positive (likes) and negative (dislikes) endorsement impacts individual belief and engagement intentions. Here, the focus shifts from message veracity to the broader influence of social endorsement on belief and engagement with online content.

To investigate how different types of social endorsement influence belief and engagement intentions, we conducted three experiments that examine the distinct and combined effects of the positive (likes) and negative (dislikes) endorsements of others. Experiment 1 isolated the separate effects of positive and negative endorsement on belief and sharing intentions. Experiment 2 replicated this design and extended the investigation to include intentions to actively like or dislike the content. Experiment 3 examined the

[1]The series of experiments reported was preceded by a pilot study conducted with undergraduate student participants, which helped refine the experiment design and materials. The data and R Notebook associated with the pilot study are available on the Open Science Framework: https://osf.io/7n3jg/

combined effects of positive and negative endorsement on belief and intentions to like, dislike, and share. By presenting the positive and negative endorsement metrics together, we tested if negative endorsement is capable of attenuating or eliminating the influence of positive endorsement by signaling that opinion is divided (Experiment 3).

Experiment 1

Message veracity has a strong effect on persuasion, but a weak effect on sharing behaviour (Fay et al., 2025). Experiment 1 provided a replication of this finding, testing the prediction that participants' belief in news content will be strongly influenced by its veracity, whereas veracity will exert a weaker effect on sharing intentions. Experiment 1 also replicated and extended prior studies showing that high levels of positive social endorsement increase belief in and intention to share false information (Avram et al., 2020; Butler et al., 2022; Luo et al., 2022; Shin, 2022; Vlasceanu & Coman, 2022). It extended this research by examining the effects of negative social endorsement. In addition, rather than focus narrowly on false information, the experiment tested how positive and negative endorsement influence belief in and engagement with information more generally (i.e., true and false information). We predicted that a high level of positive social endorsement—operationalised as a high number of likes, indicated by a "thumbs up" icon—would increase belief in and intention to share the post. Conversely, we predicted that a high level of negative social endorsement—operationalised as a high number of dislikes, indicated by a "thumbs down" icon—would reduce belief in and intention to share the content.

Using a social media simulator similar to Butler, Lamont et al. (2024), participants were shown true and false posts paired with a high or low level of positive or negative social endorsement. They then rated their belief in each post and their intention to share it. We examined positive and negative social endorsement separately to isolate their independent effects on belief and sharing intentions.

Method

Each experiment received approval from the University of Western Australia Ethics Committee. Participants viewed an information sheet before giving consent to take part in the experiment. All methods were performed in accordance with the guidelines from the National Health and Medical Research Council/Australian Research Council/University Australia's National Statement on Ethical Conduct in Human Research.

Design

The experiment used a 2 × 2 within-subjects factorial design with endorsement valence (Positive, Negative) and endorsement level (Low, High) fully crossed, plus a control condition with no endorsements. This resulted in five conditions: low positive endorsement (1–50 likes, 0 dislikes), high positive endorsement (51–100 likes, 0 dislikes), low negative endorsement (0, 1–50 dislikes), high negative endorsement (0 likes, 51–100 dislikes), and no-endorsement control (0 likes, 0 dislikes). The number of endorsements was randomly sampled from the appropriate range (including endpoints). Each participant received 40

posts in total, constituting 8 posts for each condition (4 true and 4 false). The same 40 posts were used for all participants, with allocation of post to endorsement conditions counterbalanced. In the final data from 355 participants, each post appeared 67–75 times in each condition, M = 71.0, SD = 1.5. The order in which participants received posts/conditions was randomised, but with the restriction that each quarter of the task (e.g., posts 1–10 being the first quarter) contained two posts from each condition. Distributing the conditions in this way ensured that participants had a sense of the range of endorsement possibilities from early on.

Participants

A convenience sample of 360 participants was recruited from the crowdsourcing platform Prolific (https://www.prolific.com), pre-screened to have US nationality and current US residence, and to use social media at least once per month. Testing took place on 7–8 August 2023. The most commonly used social media platforms were YouTube (324, 90.0%), Facebook (274, 76.1%), Instagram (237, 65.8%), Twitter (235, 65.3%) and Reddit (221, 61.4%). Participants were aged 19–94 (M = 38.3, SD = 12.7), with 165 (45.8%) self-identifying as women, 186 (51.7%) as men, and 8 (2.2%) as non-binary (1 preferred not to respond). Most participants (270, 75.0%) reported receiving some form of tertiary education (including 52, 14.4%, with graduate or doctoral degrees), alongside 84 (23.3%) reporting secondary education only and 3 (0.8%) reporting no formal education (3 preferred not to respond). Participants were paid £1.95 (approximately US$2.50) on completion of the experiment. Five participants were excluded from the analysis for failing attention checks.

Materials

Each participant viewed 40 mock social media posts, which took the form of news headlines paired with images (see Figure 1 for an example and Supplement 1 for the full list of the headlines used). Half of the posts were true (e.g., “Americans are using Apple AirTags to track loved ones with dementia”, “Dominican Republic starts work on border wall with Haiti”) and half were false (e.g., “Bill Clinton brought his teenage daughter to Epstein island”, “ATTENTION MEN: Iceland is giving $5000 PER MONTH to immigrants who marry Icelandic women!”). These headlines were collated from fact-checking websites, news websites, and materials used in previous studies. Headlines were selected to ensure they spanned a range of news topics, including US politics, world politics, human interest, health, and science.

Procedure

Participants began by providing their age, gender and education level, and responded to five multiple-choice questions about their social media use, adapted from past research (Brunborg & Andreas, 2019; Coyne et al., 2020; Mess et al., 2019; Scott et al., 2017). The questions concerned how much time participants spend on social media each day, how often they access social media, whether social media is part of their daily routine, how often they share news content on social media, and how often they share content in general on social media.

Participants then completed the main task. They were asked to imagine that they were using a new social media platform, and were presented with a series of 40 news headlines as social media posts on the platform (see Figure 1 for an example). The number of post likes (positive endorsement) was presented next to a thumbs up icon, and the number of dislikes (negative endorsement) next to a thumbs down icon, ostensibly based on the behaviour of past users on the social media platform. Participants were told that no number meant 0 likes or dislikes (depending on the icon). After viewing the post, participants clicked a "Next" button to respond to three questions: "How believable is the information in the above post?", "How likely would you be to share the above post online (e.g., on social media)?", and "How likely would you be to share the above post offline (e.g., during conversation)?". Responses to all three questions were given on a 5-point scale (1 = *Not at all*, 2 = *Slightly*, 3 = *Somewhat*, 4 = *Very much*, 5 = *Extremely*). Participants then clicked "Next" to proceed to the next post. As an attention check, after every ten posts (i.e., four times across the experiment) participants were asked to identify the previous page's headline from a list of four options (the correct answer plus three headlines not used in the experiment); we excluded participants with less than 75% accuracy on the attention checks. After responding to all posts, participants indicated whether they consented to their data being used for research purposes, and were then debriefed. The median completion time was 11 minutes.

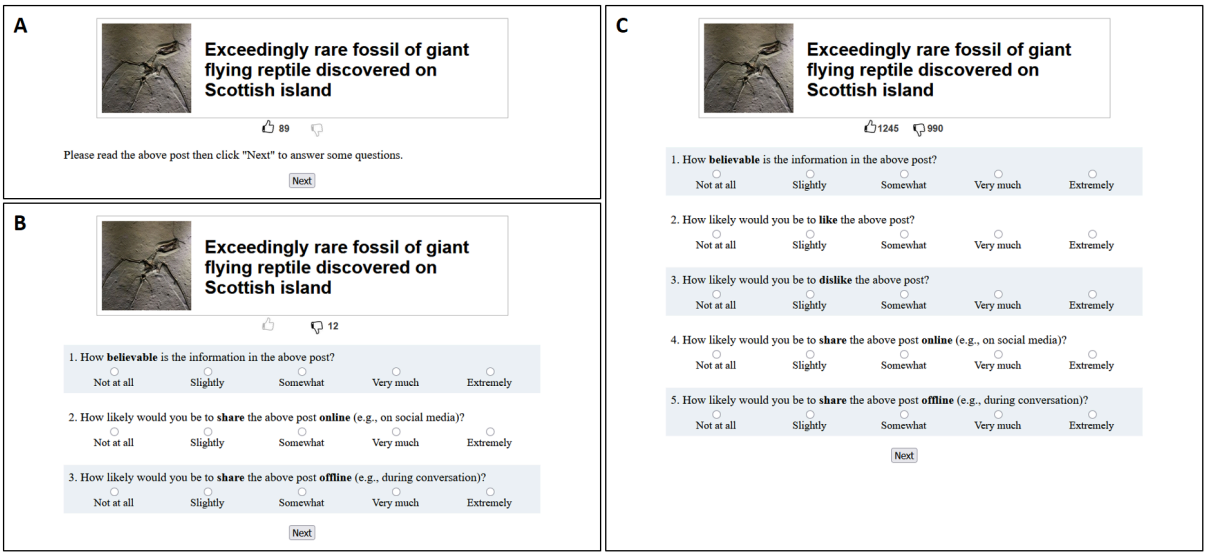

Figure 1. Screenshots from the task: (A) a post as initially shown to participants in all experiments (89 likes, 0 dislikes, as in the High Positive Endorsement condition of Experiment 1); (B) the same post after clicking "Next" in Experiment 1, with questions about belief and sharing (0 likes, 12 dislikes, as in the Low Negative Endorsement condition of Experiment 1); (C) the same post after clicking "Next" in Experiment 2 or 3, with questions about belief, liking, disliking and sharing (1245 likes, 990 dislikes, as in the High Negative Endorsement condition of Experiment 3).

Statistical Analysis

The data were analyzed using linear mixed effects modeling. For each analysis, the maximal random effects structure justified by the experiment design was specified where possible (Barr et al., 2013). All analyses were performed and all figures were created in R (R Core Team, 2013). Statistical models were estimated using the lmer() function of the lmerTest package (Bates et al., 2015; Kuznetsova et al., 2017). We also report the variance accounted for by the fixed effects using Marginal R^2. The data, R Notebooks and Supplementary Materials associated with Experiments 1–3 are provided on the Open Science Framework: https://osf.io/7n3jg/

Results

We begin by presenting descriptive statistics on participants' real-world social media use, and then examine the validity of the experimental task by testing the relationship between self-reported online sharing behaviour (i.e., in the real world) and intentions to share the news content in the task. Next, we examine the correlations between the outcomes of interest and assess participants' belief in the true and false news content, along with their intentions to share the true and false news content. Finally, we investigate the influence of positive and negative social endorsement on participants' belief in and intention to share news content.

Social Media Use and the Validity of Online Sharing Intentions

Participants reported frequently using social media, but infrequently sharing content on social media. Most spent one to two hours per day on social media, accessed social media multiple times per day and considered it a regular part of their daily routine. Despite this high usage, content sharing was rare: the most common response was never sharing content (Figure 2 Panels A–E). There was a medium-to-large correlation between self-reported online news sharing and participants' intention to share news content in the experimental task, $r(353) = .59$, $p < .001$, suggesting that sharing intentions in the experimental task serve as a robust proxy for actual online sharing behaviour.

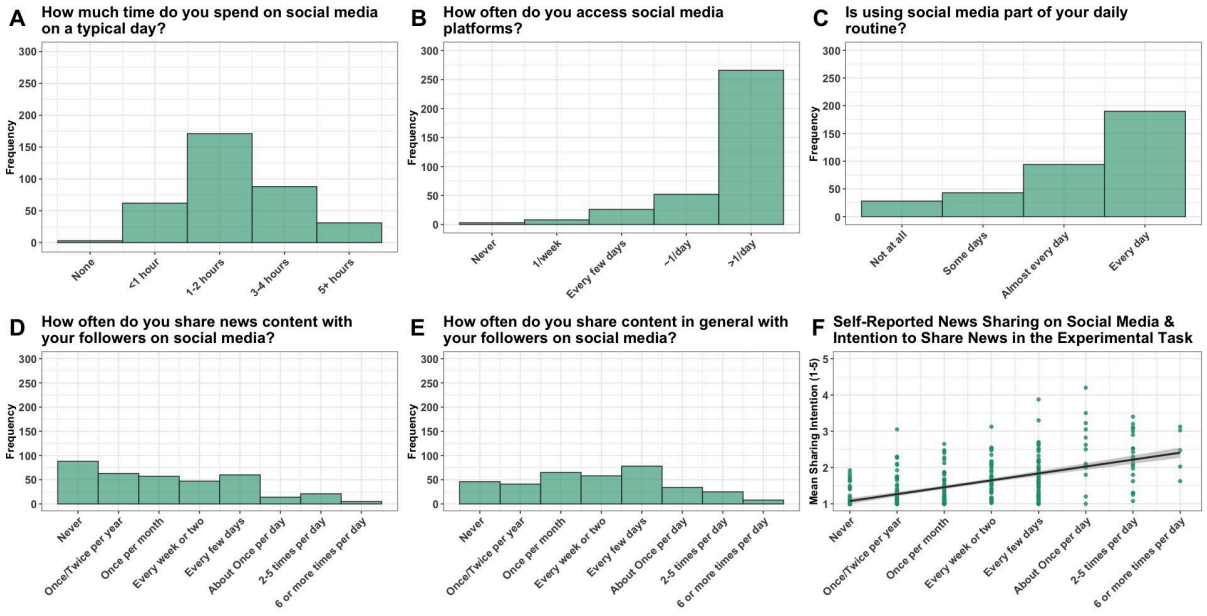

Figure 2. Self-reported measures of social media use and sharing behaviour: (A) time spent on social media per day, (B) frequency of accessing social media, (C) extent to which social media is part of a daily routine, (D) frequency of sharing news content, and (E) frequency of sharing content in general. The final panel (F) shows the relationship between self-reported online news sharing and the self-reported intention to share news in the experimental task.

Belief in and Engagement with True and False News Content

The belief and sharing outcomes were positively correlated, with correlations ranging from medium to large in size (Figure 3 Panel A; *p*s < .001). The correlations between belief and sharing intentions indicate that participants were more willing to share news content they believed to be true. Participants discriminated between true and false news content, rating the true news as more believable than the false news ($\beta = 1.06$, $t = 7.42$, $p < .001$; Marginal $R^2 = .17$). They were also more willing to share true news online and offline compared to false news ($\beta = 0.16$, $t = 2.99$, $p = .003$; Marginal $R^2 < .01$; $\beta = 0.22$, $t = 3.08$, $p = .002$; Marginal $R^2 = .01$; Figure 3 Panels B–D respectively).

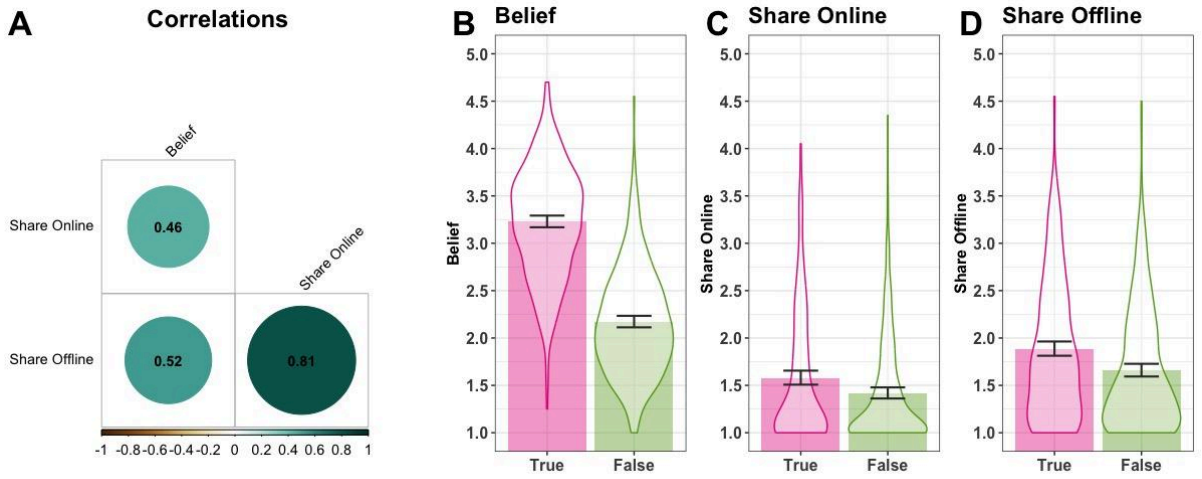

Figure 3. Correlation between outcomes (A). True versus false news: Belief (B), Share Online (C) and Share Offline (D). In panels B–D the coloured bars indicate the mean score for each condition and the violins provide distributional information. Error bars are the bootstrapped 95% CIs.

The Effect of Positive and Negative Endorsement on Belief in and Sharing of News Content

For each outcome, we first analysed the data in a 2×2 factorial design (Endorsement Valence: Positive/Negative by Endorsement Level: High/Low) that excluded the no-endorsement control condition. Any effects were then analysed in the context of the control condition. In each analysis news veracity (true, false) was entered as a covariate.

For the belief ratings, there was a significant effect of Endorsement Valence (β = 0.12, $t = 4.24$, $p < .001$), but no evidence for an effect of Endorsement Level ($p = .376$) or interaction between Valence and Level ($p = .855$; Marginal $R^2 < .01$). The effect of Valence indicates that positively endorsed news content was believed more than negatively endorsed content. Whereas positive endorsement increased belief relative to the no-endorsement control condition (β = 0.06, $t = 2.65$, $p = .008$), negative endorsement decreased belief relative to the no-endorsement control condition (β = −0.06, $t = -2.45$, $p = .014$; Figure 7 Panel A).

For the online sharing ratings, there was a significant effect of Endorsement Valence (β = 0.06, $t = 3.25$, $p = .001$), but no evidence for an effect of Endorsement Level ($p = .508$) or interaction between Valence and Level ($p = .465$; Marginal $R^2 < .01$). The effect of Valence indicates that participants were more willing to share positively endorsed news content online than negatively endorsed news content. Positive endorsement increased online sharing intentions relative to the no-endorsement control condition (β = 0.05, $t = 2.80$, $p = .008$), but negative endorsement did not differ from the control ($p = .711$; Figure 7 Panel B). Unsurprisingly, given the high correlation between online and offline sharing intentions ($r = .81$), the same pattern of results was observed for offline sharing (Figure 7 Panel C).

The Experiment 1 findings show that news veracity had a large effect on belief, explaining over 17% of the variance in belief ratings. News veracity also influenced online and offline sharing intentions, but the effect was small, explaining 1% or less of the variance

in sharing ratings. This asymmetry between believing and sharing replicates Fay et al. (2025). Positive social endorsement increased belief in and willingness to share the news content. Negative social endorsement reduced belief in the content, but had no detectable effect on sharing intentions. The variance explained by the endorsement effects was small, accounting for less than 1% of the outcome variance.

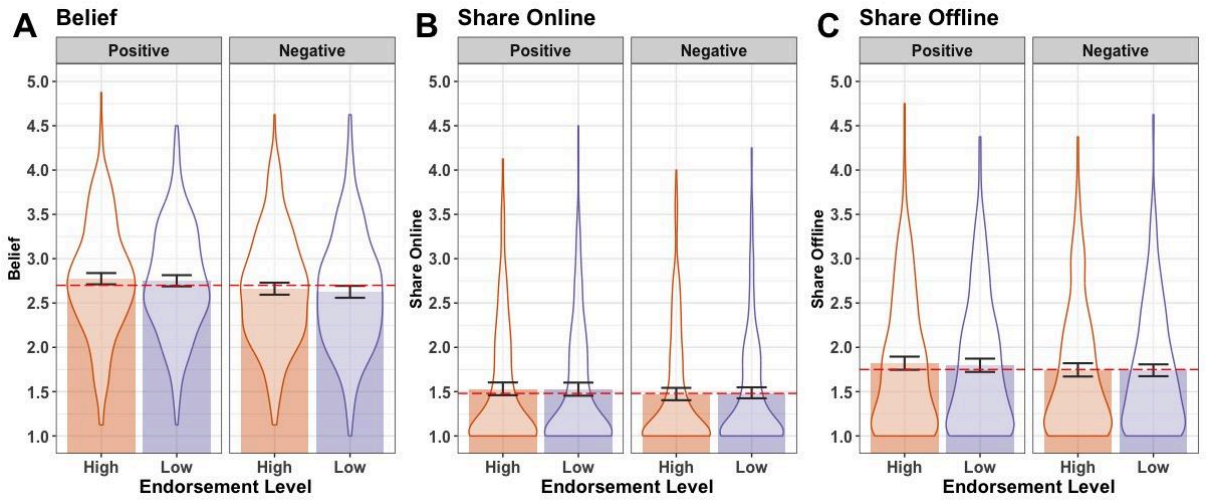

Figure 4. The effect of high and low levels of positive and negative social endorsement on: Belief (A), Sharing Online (B) and Sharing Offline (C). The red dashed horizontal line conveys the mean score for the no-endorsement control condition. The coloured bars indicate the mean score for each condition and the violins provide distributional information. Error bars are the bootstrapped 95% CIs.

Experiment 2

Experiment 2 replicated the design of Experiment 1 but also measured participants' intentions to like or dislike the content. This allowed us to examine how people use positive and negative social endorsements to respond to true and false news, and how others' endorsements directly influence participants' own endorsement intentions, as well as their belief in and intention to share the content. We also modified the level of social endorsement, such that the number of likes/dislikes across Endorsement Levels (high vs. low) was more pronounced than in Experiment 1, consistent with Butler et al. (2022) and more realistic for some platforms (especially for "viral" posts). We reasoned that amplifying the difference between high and low levels of social endorsement would strengthen the salience of the manipulation, making any effects on belief, like/dislike and sharing intentions easier to detect.

Method

Design

Experiment 2 used the same design same as Experiment 1, but the number of endorsements was changed so that the low number of endorsements was sampled from a normal distribution with M = 10 and SD = 4 (instead of ranging 1–50), and the high number of

endorsements was sampled from a normal distribution with *M* = 1000 and *SD* = 200 (instead of ranging 51–100). In both cases the normal distribution was truncated so the minimum number of endorsements (when intended to be more than 0) was 1. In the final data from 359 participants, each post appeared 68–76 times in each condition, *M* = 71.8, *SD* = 1.6.

Participants

A convenience sample of 362 participants was recruited from Prolific, pre-screened (as per Experiment 2) to have US nationality and current US residence, and to use social media at least once per month. Participants from Experiment 1 were not permitted to take part in Experiment 2. Testing took place on 28–29 September 2023. The most commonly used social media platforms were YouTube (330, 91.2%), Facebook (265, 73.2%), Instagram (239, 66.0%), Twitter (211, 58.3%) and Reddit (208, 57.5%). Participants were aged 18–76 (*M* = 41.1, *SD* = 14.0), with 183 (50.6%) self-identifying as women, 174 (48.1%) as men, and 4 (1.1%) as non-binary (1 preferred not to respond). Most participants (266, 73.5%) reported receiving some form of tertiary education (including 57, 15.7%, with graduate or doctoral degrees), alongside 89 (24.6%) reporting secondary education only and 5 (1.4%) reporting no formal education (2 preferred not to respond). Participants were paid £1.95 (approximately US$2.50) on completion of the experiment. Three participants were excluded from analysis for failing attention checks.

Materials and Procedure

The materials and procedure were the same as in Experiment 1 except that participants were asked two additional questions for each post: “How likely would you be to like the above post?” and “How likely would you be to dislike the above post?”. These questions used the same 5-point response scale as the other questions. The median completion time was 15 minutes.

Results

Social Media Use and the Validity of Online Sharing Intentions

Like the Experiment 1 participants, the Experiment 2 participants frequently used social media but rarely shared content (Figure 5 Panels A–E). There was a medium-to-large correlation between self-reported online news sharing and intention to share news content in the experimental task, $r(353) = .57$, $p < .001$, suggesting that sharing intentions in the experimental task serve as a robust proxy for actual online sharing behaviour.

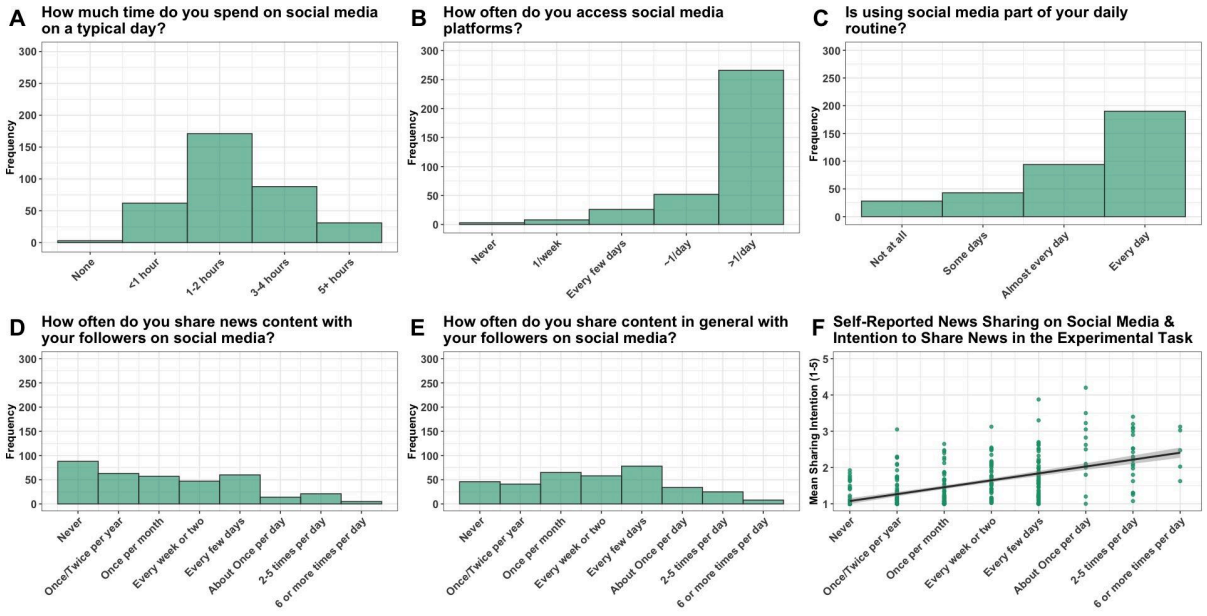

Figure 5. Self-reported measures of social media use and sharing behaviour: (A) time spent on social media per day, (B) frequency of social media access, (C) extent to which social media is part of the daily routine, (D) frequency of sharing news content, and (E) frequency of sharing content in general. The final panel (F) shows the relationship between self-reported online news sharing and the self-reported intention to share news in the experimental task.

Belief in and Engagement with True and False News Content

The belief, like/dislike, and sharing outcomes were positively correlated, with correlations ranging from medium to large in size (Figure 6, Panel A; *p*s < .001). As per Experiment 1, the correlations between belief and sharing intentions indicate that participants were more willing to share news content they believed to be true. The correlations between belief and liking/disliking indicate that participants were more willing to like or dislike content they believed to be true, although this relationship was stronger for liking. A similar pattern is observed for sharing.

Participants discriminated between true and false news content, rating true news as more believable than false news (β = 1.02, t = 7.35, p < .001; Marginal R^2 = .15). They were also more willing to like true news (β = 0.18, t = 2.57, p = .010; Marginal R^2 = .01), dislike false news (β = −0.17, t = −2.54, p = .011; Marginal R^2 < .01), share true news online (β = 0.09, t = 2.19, p = .028; Marginal R^2 < .01) and share true news offline (β = 0.13, t = 2.07, p = .039; Marginal R^2 < .01; Figure 6 Panels B–F respectively).

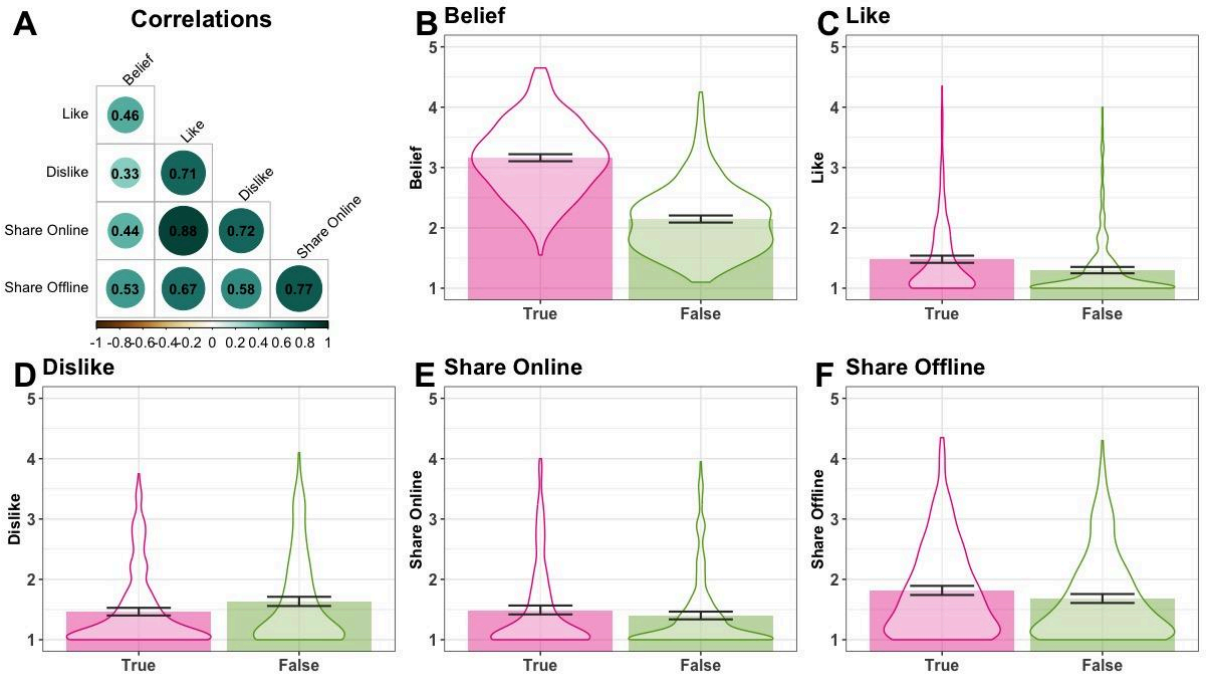

Figure 6. Correlation between outcomes (A). True versus false news: Belief (B), Like (C), Dislike (D), Share Online (E) and Share Offline (F). In panels B–F the coloured bars indicate the mean score for each condition and the violins provide distributional information. Error bars are the bootstrapped 95% CIs.

The Effect of Positive and Negative Endorsement on Belief in and Engagement with News Content

For the belief ratings, there was a significant effect of Endorsement Valence (β = 0.11, $t = 3.63$, $p < .001$), but no evidence for an effect of Endorsement Level ($p = .905$) or interaction between Valence and Level ($p = .099$; Marginal $R^2 < .01$). The effect of Valence indicates that positively endorsed news content was believed more than negatively endorsed news content. Whereas positive endorsement increased belief relative to the no-endorsement control condition (β = 0.06, $t = 2.39$, $p = .017$), belief ratings in the negative endorsement condition did not differ from the control ($p = .239$; Figure 7 Panel A).

For the like ratings, there was a significant effect of Endorsement Valence (β = 0.14, $t = 6.36$, $p < .001$), no evidence for an effect of Endorsement Level ($p = .660$) and a significant interaction between Valence and Level (β = −0.06, $t = -2.37$, $p = .018$; Marginal $R^2 < .01$). The effect of Valence indicates that participants were more willing to like positively endorsed news content than negatively endorsed news content. The interaction was driven by a significant effect of Endorsement Level in the Positive Endorsement condition (High > Low; $p = .005$), but not in the Negative Endorsement condition ($p = .635$). High and low levels of positive social endorsement increased participants' intention to like the news content relative to the no-endorsement control condition ($ps < .007$). High levels of negative endorsement decreased participants' intention to like the news content relative to the no-endorsement control ($p = .028$), but low levels of negative endorsement did not ($p = .085$;

Figure 7 Panel B). The dislike ratings returned a similar pattern of results: a significant main effect of Endorsement Valence ($\beta = -0.19$, $t = -5.91$, $p < .001$) and Endorsement Level ($\beta = -0.07$, $t = -3.11$, $p = .002$) and an interaction between Valence and Level ($\beta = 0.07$, $t = 2.12$, $p = .034$; Marginal $R^2 < .01$). The effect of Valence indicates that negatively endorsed news content increased participants' intention to dislike the news content relative to positively endorsed news content. The interaction was driven by a significant effect of Endorsement Level in the Negative Endorsement condition (High > Low; $p = .003$), but not in the Positive Endorsement condition ($p = .904$). High and low levels of negative social endorsement increased participants' intention to dislike the news content relative to the no-endorsement control condition (ps < .039). By contrast, high and low levels of positive endorsement decreased participants' intention to dislike the content relative to the no-endorsement control ($p = .001$; Figure 7 Panel C).

For the online sharing ratings, there was a significant effect of Endorsement Valence ($\beta = 0.06$, $t = 3.21$, $p = .001$), but no evidence for an effect of Endorsement Level ($p = .498$) or interaction between Valence and Level ($p = .465$; Marginal $R^2 < .01$). The effect of Valence indicates that participants were more willing to share positively endorsed news content online than negatively endorsed news content. Positive endorsement increased online sharing intentions relative to the no-endorsement control condition ($\beta = 0.03$, $t = 2.15$, $p = .032$), but negative endorsement did not differ from the control ($p = .331$; Figure 7 Panel D). For offline sharing, there was a significant effect of Endorsement Valence ($\beta = 0.10$, $t = 4.56$, $p < .001$), no evidence for an effect of Endorsement Level ($p = .998$) and a significant interaction between Valence and Level ($\beta = -0.07$, $t = -2.19$, $p = .029$; Marginal $R^2 < .01$). The effect of Valence indicates that participants were more willing to share positively endorsed news content offline than negatively endorsed news content. The interaction was driven by a significant effect of Endorsement Level in the Positive Endorsement condition (High > Low; $p = .002$), but not in the Negative Endorsement condition ($p > .999$). High levels of positive social endorsement increased offline sharing intentions relative to the no-endorsement control condition ($p = .002$), but low levels of positive endorsement did not ($p = .989$). Offline sharing intentions in the negative endorsement conditions (high and low) did not differ from the control (ps > .101; Figure 7 Panel E).

Replicating Experiment 1, news veracity had a large effect on belief, accounting for over 15% of the variance in belief ratings. By contrast, its effect on online and offline sharing intentions was small, explaining less than 1% of the variance. Veracity also influenced endorsement intentions: participants were more willing to like true news and dislike false news. These effects were also small, explaining 1% or less of the variance in endorsement ratings.

As in Experiment 1, positive social endorsement increased belief in the news content. Contrary to Experiment 1, negative endorsement had no detectable effect on belief. Participants were more willing to like content that received positive endorsement, particularly when the endorsement level was high. Conversely, high levels of negative endorsement decreased participants' intention to like the content. A complementary

pattern was observed for disliking: participants were more willing to dislike content with negative endorsement, especially at high levels, and less willing to dislike content with positive endorsement. Replicating Experiment 1, positive social endorsement increased online and offline sharing intentions, with high levels of positive endorsement further boosting offline sharing. Negative social endorsement, once again, had no detectable effect on sharing intentions. Across outcomes, the effects of social endorsement were small, accounting for less than 1% of the variance.

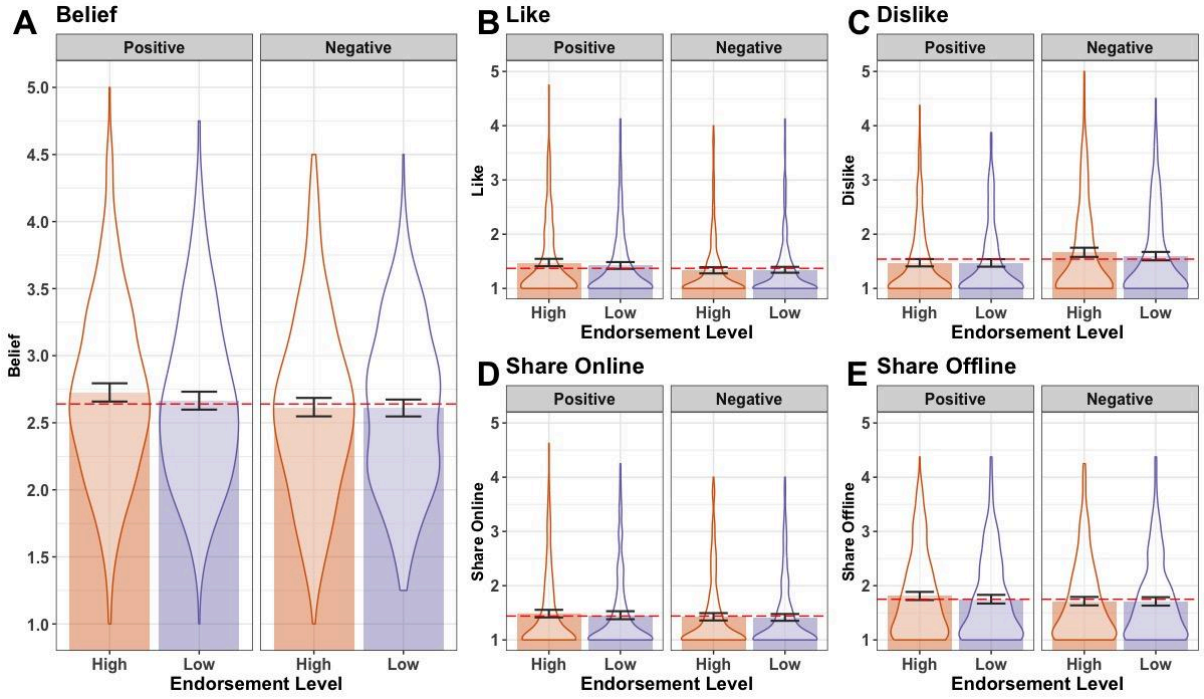

Figure 7. The effect of high and low levels of positive and negative social endorsement on: Belief (A), Liking (B), Disliking (C), Sharing Online (D) and Sharing Offline (E). The red dashed horizontal line conveys the mean score for the no-endorsement control condition. The coloured bars indicate the mean score for each condition and the violins provide distributional information. Error bars are the bootstrapped 95% CIs.

Experiment 3

Experiment 3 examined the consequences of contested social endorsement—that is, news posts that received both positive and negative social endorsement. While people often conform to majority opinions, research shows that the presence of even a single dissenter can significantly reduce conformity to the majority (Allen & Levine, 1969; Asch, 1955). In Experiments 1–2, positive social endorsement consistently increased belief in and intention to share posts, whereas the effect of negative endorsement was inconsistent. Experiment 3 tested whether negative social endorsement exerts more influence not in isolation, but by attenuating the impact of positive endorsement. To examine this, belief, like/dislike and sharing intentions were compared across four conditions: (1) high positive endorsement only, representing an uncontested opinion condition, (2) high positive endorsement with low

negative endorsement, representing a contested opinion condition, (3) high positive endorsement with high negative endorsement, representing a highly contested opinion condition, and (4) a no-endorsement control condition.

Method

Design

Experiment 3 had four within-subjects conditions (high positive endorsement only, high positive and low negative endorsement, high positive and high negative endorsement, no-endorsement control). Participants received 10 posts in each condition. In the final data from 354 participants, each post appeared 83–93 times in each condition, M = 88.5, SD = 1.9. The order in which participants received posts/conditions was randomised with a similar restriction to that in Experiments 1–2, whereby each fifth of the task (e.g., posts 1–8 being the first fifth) contained two posts from each condition. The number of endorsements corresponding to the high and low conditions was calculated as per Experiment 2.

Participants

A convenience sample of 362 participants was recruited from Prolific, pre-screened (as per Experiments 1 and 2) to have US nationality and current US residence, and to use social media at least once per month. Participants from Experiments 1 and 2 were not permitted to take part in Experiment 3. Testing took place on 18–21 March 2024. The most commonly used social media platforms were YouTube (324, 89.5%), Facebook (291, 80.4%), Instagram (238, 65.7%), Twitter (211, 58.3%) and Reddit (211, 58.3%). Participants were aged 20–75 (M = 40.9, SD = 12.2), with 184 (50.8%) self-identifying as women, 173 (47.8%) as men, and 3 (0.8%) as non-binary (2 preferred not to respond). Most participants (280, 77.3%) reported receiving some form of tertiary education (including 67, 18.5%, with graduate or doctoral degrees), alongside 80 (22.1%) reporting secondary education only (2 preferred not to respond). Participants were paid £1.95 (approximately US$2.50) on completion of the experiment. Eight participants were excluded from analysis for failing attention checks.

Materials and Procedure

The materials and procedure were the same as in Experiment 2. The median completion time was 15 minutes.

Results

Social Media Use and the Validity of Online Sharing Intentions

As per Experiments 1–2, the Experiment 3 participants frequently consumed content on social media but rarely shared content (Figure 8 Panels A–E). There was a medium-to-large correlation between self-reported online news sharing and intention to share news in the experimental task, $r(352) = .47$, $p < .001$, suggesting that sharing intentions in the experimental task serve as a robust proxy for actual online sharing behaviour.

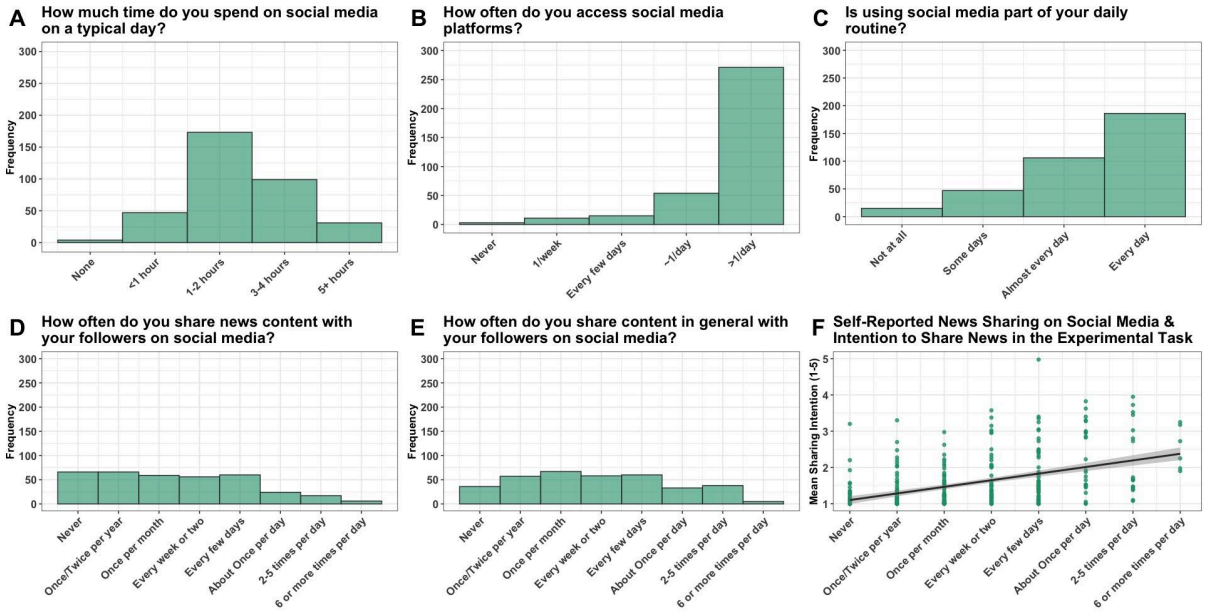

Figure 8. Self-reported measures of social media use and sharing behaviour: (A) time spent on social media per day, (B) frequency of social media access, (C) extent to which social media is part of the daily routine, (D) frequency of sharing news content, and (E) frequency of sharing content in general. The final panel (F) shows the relationship between self-reported online news sharing and the self-reported intention to share news in the experimental task.

Belief in and Engagement with True and False News Content

The belief, like/dislike, and sharing outcomes were positively correlated, with correlations ranging from medium to large in size (Figure 9 Panel A; *p*s < .001). Replicating Experiment 1 and 2, the correlations between belief and sharing intentions indicate that participants were more willing to share news content they believed to be true. Replicating Experiment 2, the correlations between belief and liking/disliking indicate that participants were more willing to like or dislike content they believed to be true, although this relationship was stronger for liking. A similar pattern was again observed for sharing.

Participants discriminated between true and false news content, rating true news as more believable than false news content (β = 0.97, *t* = 6.97, *p* < .001; Marginal R^2 = .14). They were also more likely to like true news (β = 0.21, *t* = 2.58, *p* = .010; Marginal R^2 = .01), dislike false news (β = −0.22, *t* = −2.85, *p* = .004; Marginal R^2 < .01) and share true news online (β = 0.10, *t* = 2.19, *p* = .028; Marginal R^2 < .01), but not offline, although this effect was marginal (*p* = .070; Figure 9 Panels B–F respectively). These results replicate those from Experiment 1 and 2.

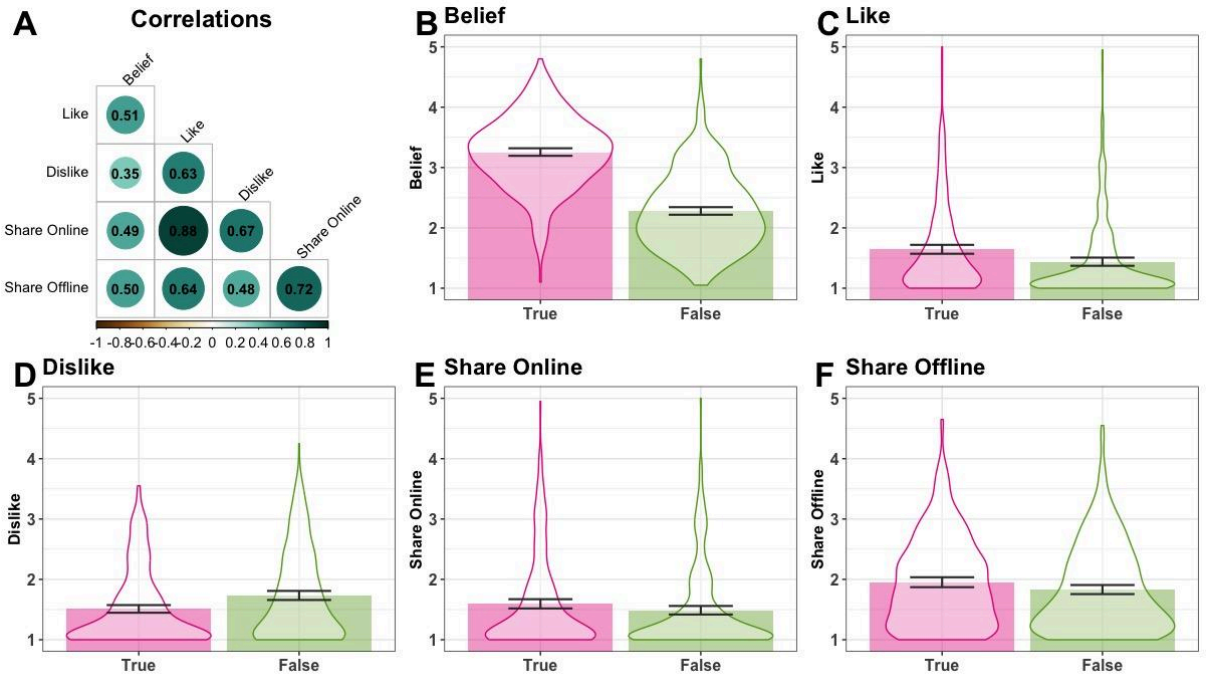

Figure 9. Correlation between outcomes (A). True versus false news: Belief (B), Like (C), Dislike (D), Share Online (E) and Share Offline (F). In panels B–F the coloured bars indicate the mean score for each condition and the violins provide distributional information. Error bars are the bootstrapped 95% CIs.

The Effect of Combining Positive and Negative Endorsement on Belief in and Engagement with News Content

For the belief ratings, news content associated with exclusively high positive endorsement was believed more than news content with high positive and high negative social endorsement ($\beta = -0.07$, $t = -2.92$, $p = .004$; Marginal $R^2 < .01$). News content with high positive endorsement and low negative endorsement was rated to be as believable as news content with exclusively high positive endorsement ($p = .803$). There was no statistical evidence of a difference between the experimental conditions and the no-endorsement control condition ($ps > .105$; Figure 10 Panel A).

For the like ratings, participants were more willing to like news content with exclusively high positive endorsement than content with high positive and high negative endorsement ($\beta = -0.07$, $t = -3.90$, $p < .001$; Marginal $R^2 < .01$). Participants were similarly willing to like news content with exclusively high positive endorsement and content with high positive and low negative endorsement ($p = .862$). Participants were more willing to like news content in the exclusively high positive endorsement condition and the high positive with low negative endorsement condition compared to the no-endorsement control condition ($\beta = 0.05$, $t = 2.43$, $p = .015$; $\beta = 0.05$, $t = 2.60$, $p = .009$ respectively; Marginal $R^2 = .01$). There was no statistical evidence of a difference between the high positive and high negative endorsement condition and the no-endorsement control condition ($p = .141$; Figure 10 Panel B). The opposite pattern was observed for the dislike ratings. Here, participants

were more willing to dislike news content with high positive and high negative endorsement compared to news content with exclusively high positive endorsement (β = 0.13, t = 5.88, p < .001; Marginal R^2 < .01). Participants were similarly willing to dislike news content with exclusively high positive endorsement and content with high positive and low negative endorsement (p = .165). Participants were more willing to dislike news content in the high positive and high negative endorsement condition compared to the no-endorsement control condition (β = 0.10, t = 4.65, p < .001; Marginal R^2 = .01). There was no statistical evidence of a difference between either the exclusively high positive endorsement condition or the high positive and high negative endorsement condition when compared to the no-endorsement control condition (ps > .213; Figure 10 Panel C).

For the online sharing ratings, there was no statistical evidence of a difference between the experimental conditions, or between the experimental conditions and the no-endorsement control condition (ps > .055; Figure 10 Panel D). The same pattern of null results was observed for the offline sharing ratings (ps > .261; Figure 10 Panel E).

Replicating Experiment 1 and 2, news veracity had a large effect on belief, accounting for over 14% of the variance in belief ratings. Veracity also influenced endorsement intentions: participants were more willing to like true news and dislike false news, replicating Experiment 2. Again, the effects were small, explaining 1% or less of the variance in endorsement ratings. Replicating Experiment 1 and 2, veracity influenced online and offline sharing intentions, again with a small effect that accounted for less than 1% of the variance. Contrary to Experiment 1 and 2, there was no statistical evidence that news veracity affected offline sharing.

The primary purpose of Experiment 3 was to examine the effects of contested social endorsement. News content paired with exclusively high positive endorsement was believed more than content paired with both high positive and high negative endorsement. However, introducing a low level of negative endorsement alongside high positive endorsement did not reduce belief. A similar pattern emerged for liking intentions: participants were more willing to like content with exclusively high positive endorsement than content paired with both high positive and high negative endorsement. Including a low level of negative endorsement did not reduce participants' intention to like the content. For disliking, the pattern was reversed: participants were more willing to dislike content paired with high positive and high negative endorsement than exclusively high positive endorsement. Again, including a low level of negative endorsement did not increase participants' intention to dislike the content. In each case, the effects were small, accounting for 1% or less of the variance in the outcomes. Unlike Experiments 1 and 2, positive social endorsement did not increase participants' willingness to share the news content online or offline—in Experiment 3, sharing intentions did not differ across endorsement conditions or relative to the no-endorsement control.

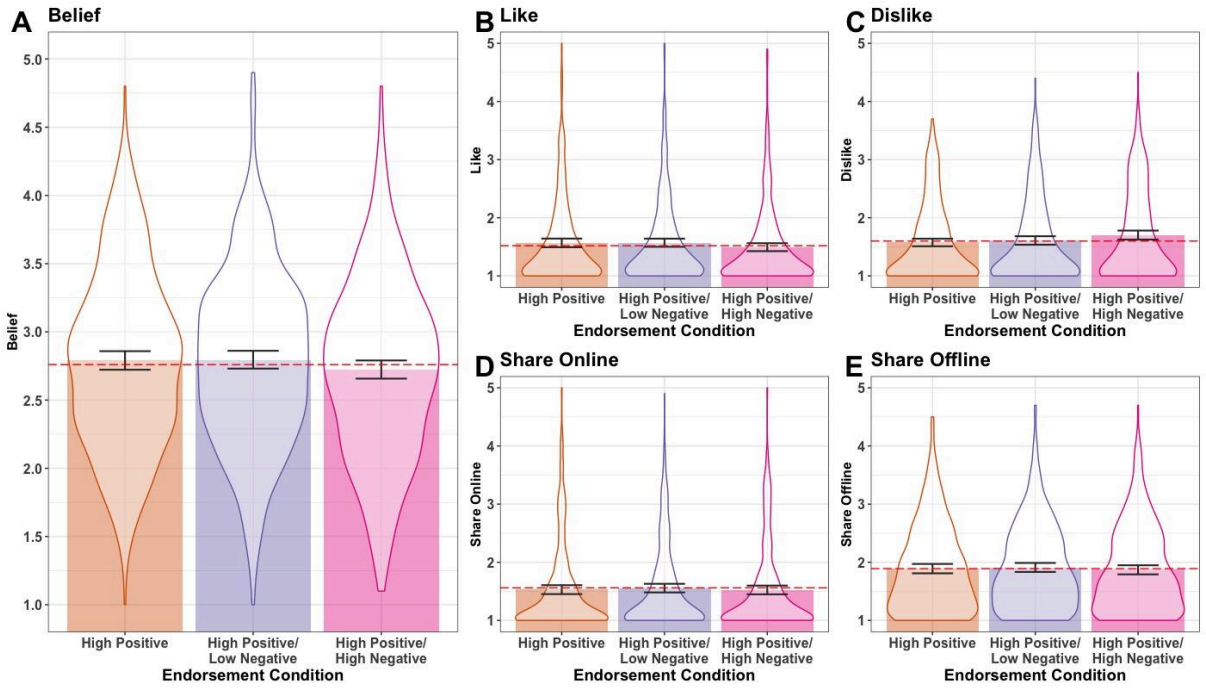

Figure 10. The effect of combining positive and negative (high or low) social endorsement on: Belief (A), Liking (B), Disliking (C), Sharing Online (D) and Sharing Offline (E). The red dashed horizontal line conveys the mean score for the no-endorsement control condition. The coloured bars indicate the mean score for each condition and the violins provide distributional information. Error bars are the bootstrapped 95% CIs.

Discussion

The experiments reported make three distinct contributions: 1) an examination of people's social media use, 2) how people evaluate and engage with true and false online news content, and 3) how others' positive and negative social endorsements impact individual belief, endorsement intentions, and sharing behaviour.

Across all three experiments, participants reported high levels of social media use but low levels of content sharing. Most participants accessed social media multiple times a day and regarded it as an established part of their daily routine, yet the most common response was never sharing content. This highlights a discrepancy between consumption and production on social media: users frequently engage as passive consumers rather than active disseminators of content. This is consistent with the finding that 75–90% of people on social media are passive users (known as *lurkers*; Antelmi et al., 2019; Baqir et al., 2025; Gong et al., 2015). Self-reported sharing frequency was moderately to strongly correlated with participants' sharing intentions in the experimental task (*r*s = .47–.59), indicating the task provided a valid proxy for real-world sharing behaviour. This supports the ecological validity of our experimental measure and suggests that, despite the low base rate of sharing,

the task captured meaningful individual differences in participants' propensity to share content online.

In each experiment, participants reliably distinguished between true and false news content, rating true news as more believable than false news and showing greater willingness to engage with it. The engagement measures—like, dislike, and share—tended to be strongly correlated (*r*s = .48–.88), replicating Tenenboim (2022), and indicating a general engagement factor. Belief was positively correlated with each engagement measure, with correlations ranging from medium to large in size (*r*s = .47–.59), indicating that participants were more inclined to endorse—positively or negatively—and share content they believed to be true and not engage with content they regarded as false. This relationship was especially pronounced for liking, suggesting that belief functions as a stronger driver of positive endorsement than negative endorsement. While the effects of perceived news veracity on engagement were small, they were robust across experiments, replicating prior findings that veracity has a large influence on belief (Marginal R^2 = .14–.17) and a weaker influence on sharing behaviour (Marginal R^2 < .01) (Fay et al., 2025; see also Pfänder & Altay, 2025). These findings align with the literature on epistemic vigilance (Sperber et al., 2010) and accuracy-motivated sharing (Howe et al., 2024; Pennycook & Rand, 2022; Rathje et al., 2023), which propose that people selectively attend to content veracity and are more likely to share information they consider credible. Together, these findings underscore the role of belief as a key psychological mediator of online engagement.

The final contribution is to simulate how others' positive and negative social endorsement might guide individual beliefs and engagement with online content. In each experiment, social endorsement consistently influenced participants' belief in and engagement with online news content, though the effects were small. Positive endorsement increased belief, like and share intentions and decreased dislike intentions. By contrast, negative endorsement reduced belief (Experiment 1, but not Experiment 2) and like intentions and increased dislike intentions, but had no detectable impact on share intentions. Experiment 3 extended these findings by examining contested endorsement, showing that the presence of high negative endorsement alongside high positive endorsement reduced belief and like intentions relative to exclusively high positive endorsement (or high positive endorsement alongside low negative endorsement). A complementary pattern was observed for dislike intentions, with participants more willing to dislike content under high contested endorsement. Across all experiments, the variance explained by the endorsement effects was consistently small (Marginal R^2 < 1%). Our results demonstrate that social endorsement functions as a subtle but reliable moderator of belief and engagement, with positive cues generally enhancing acceptance and endorsement of content, replicating past research (Avram et al., 2020; Butler et al., 2022; Jucks & Thon, 2017; Luo et al., 2022; Shin et al., 2022). By contrast, negative cues were capable of dampening belief and eliciting disapproval when sufficiently salient (see also Butler, Fay, et al., 2024).

Taken together, our findings highlight that while social endorsement influences belief and engagement with online content, its effects are consistently small relative to the impact of veracity. In isolation, this suggests the impact of online information is governed less by social consensus signals than by epistemic vigilance, with belief serving as the key factor linking endorsement and dissemination. However, the limited effects of social endorsement at the individual level must be considered in the context of recommender systems which play a central role in determining what online content is visible (Jannach et al., 2010). Virtually every social media platform employs a recommender system algorithm that decides what content each user is presented with, and these algorithms are optimised to maximise content engagement and consumption (Germano et al., 2022; Narayan, 2023). Content that is popular right now (measured for example by the total number of likes or by the like-dislike ratio) will be more likely to be shown in the future while heavily disliked content may be hidden or not actively promoted, creating a "rich-get-richer" feedback loop. In addition, a user is more likely to be shown content that is popular among other users fitting a similar profile. The findings of this paper may be more consequential once recommender systems are taken into account. Even small behavioural differences in liking, disliking, or sharing can become substantial as, iteratively, the recommender system presents users with increasingly personalised and likeable content and then takes the content engagement as information to be fed back into the algorithm. Modest endorsement effects may scale to yield substantial shifts in content visibility and diffusion when filtered through recommender systems. Recently, mathematical models have been proposed to study the closed-loop between recommender systems on user beliefs and engagement, and our findings can help to improve these models (Davidson & Ye, 2025; Rossi et al., 2022).

Our results highlight that dislikes serve different psychological functions depending on the type of content being evaluated. In the experiments reported, participants used dislikes in response to bad news (e.g., reports of harmful events), with a medium effect size (*r*s = .33–.35), as well as to reject content they considered untrue, where the effect was small (Marginal $R^2 < .01$). This dual role highlights that negative endorsement can signal both a reaction to undesirable information and epistemic rejection. To clearly distinguish between these functions—and to better signal when content may be false—it may be valuable for social media platforms to pair negative endorsement with community annotation features that clarify whether a post is being disliked because it is upsetting, misleading, or false (see Bond & Garretta, 2023; Drolsbach et al., 2024; Slaughter et al., 2025; Wirtschafter & Majumder, 2023). Nonetheless, negative endorsement may still have larger cumulative effects on the sharing of false information than our results show. Participants within our experiments did not have to worry about negative endorsement (i.e., dislikes) of posts they shared. However, if dislikes were implemented then the potential for receiving negative social feedback after sharing false information may have stronger cumulative effects on behavior (e.g., Altay et al., 2022; Prike et al., 2024). When applied outside the domain of news, the dislike button may also provide a straightforward

mechanism for signalling social disapproval of individuals rather than content. In this way, it risks being co-opted as a tool for online bullying, enabling collective targeting or ostracism through aggregated negative feedback (Ray et al., 2024). Thus, while dislikes can serve a functional role in regulating the credibility of information, their potential to facilitate harmful interpersonal dynamics underscores the need to carefully weigh their broader social impact. Future research is required to establish how negative endorsement might be used in practice, and whether it poses risks of cyberbullying.

Conclusion

Across three experiments, we show that people's responses to online news are shaped most strongly by content veracity, with social endorsements exerting small but consistent additional effects. These findings extend prior work by demonstrating that belief functions as a central mediator of online engagement, while also clarifying the subtle yet reliable role of social cues. Importantly, because recommender systems amplify engagement signals, even modest differences in liking, disliking, or sharing can accumulate into substantial shifts in visibility and diffusion. Our results highlight the distinctive role of dislikes, which capture both epistemic rejection and moral condemnation. While this dual function suggests that a dislike option could enrich users' evaluative toolkit and improve content curation, it also raises the risk of misuse as a tool for online bullying.

References

Allen, V. L., & Levine, J. M. (1969). Consensus and conformity. *Journal of Experimental Social Psychology*, *5*(4), 389–399. https://doi.org/10.1016/0022-1031(69)90032-8

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

Antelmi, A., Malandrino, D., & Scarano, V. (2019). Characterizing the Behavioral Evolution of Twitter Users and The Truth Behind the 90-9-1 Rule. *Companion Proceedings of The 2019 World Wide Web Conference*, 1035–1038. https://doi.org/10.1145/3308560.3316705

Asch, S. E. (1951). Effects of group pressure upon the modification and distortion of judgments. *Organizational Influence Processes*, *58*, 295–303.

Asch, S. E. (1955). Opinions and Social Pressure. *Scientific American*, *193*(5), 31–35.

Avram, M., Micallef, N., Patil, S., & Menczer, F. (2020). Exposure to social engagement metrics increases vulnerability to misinformation. *Harvard Kennedy School Misinformation Review*. https://doi.org/10.37016/mr-2020-033

Baqir, A., Chen, Y., Diaz-Diaz, F., Kiyak, S., Louf, T., Morini, V., Pansanella, V., Torricelli, M., & Galeazzi, A. (2025). Unveiling the drivers of active participation in social media discourse. *Scientific Reports*, *15*(1), 4906. https://doi.org/10.1038/s41598-025-88117-x

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

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models

Using lme4. *Journal of Statistical Software*, *67*(1), Article 1. https://doi.org/10.18637/jss.v067.i01

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

Bebbington, K., MacLeod, C., Ellison, T. M., & Fay, N. (2017). The sky is falling: Evidence of a negativity bias in the social transmission of information. *Evolution and Human Behavior*, *38*(1), 92–101. https://doi.org/10.1016/j.evolhumbehav.2016.07.004

Bindra, S., Sharma, D., Parameswar, N., Dhir, S., & Paul, J. (2022). Bandwagon effect revisited: A systematic review to develop future research agenda. *Journal of Business Research*, *143*, 305–317. https://doi.org/10.1016/j.jbusres.2022.01.085

Bond, R. M., & Garretta, R. K. (2023). Engagement with fact-checked posts on Reddit. *PNAS Nexus*, pgad018. https://doi.org/10.1093/pnasnexus/pgad018

Boyd, R., & Richerson, P. J. (1985). *Culture and the evolutionary process*. University of Chicago Press.

Brunborg, G. S., & Andreas, J. B. (2019). Increase in time spent on social media is associated with modest increase in depression, conduct problems, and episodic heavy drinking. *Journal of Adolescence*, *74*, 201–209. https://doi.org/10.1016/j.adolescence.2019.06.013

Butler, L. H., Fay, N., & Ecker, U. K. H. (2022). Social Endorsement Influences the Continued Belief in Corrected Misinformation. *Journal of Applied Research in Memory and Cognition*. Scopus. https://doi.org/10.1037/mac0000080

Butler, L. H., Fay, N., & Ecker, U. K. H. (2024). Others (dis-)endorse this so it must (not) be true: High relative endorsement increases perceived misinformation veracity but not

correction effectiveness. *Applied Cognitive Psychology*, *38*(1), e4146. https://doi.org/10.1002/acp.4146

Butler, L. H., Lamont, P., Wan, D. L. Y., Prike, T., Nasim, M., Walker, B., Fay, N., & Ecker, U. K. H. (2024). The (Mis)Information Game: A social media simulator. *Behavior Research Methods*. https://doi.org/10.3758/s13428-023-02153-x

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), e2023301118. https://doi.org/10.1073/pnas.2023301118

Coyne, S. M., Rogers, A. A., Zurcher, J. D., Stockdale, L., & Booth, M. (2020). Does time spent using social media impact mental health?: An eight year longitudinal study. *Computers in Human Behavior*, *104*, 106160. https://doi.org/10.1016/j.chb.2019.106160

Davidson, E. C., & Ye, M. (2025). *Modelling the Closed Loop Dynamics Between a Social Media Recommender System and Users' Opinions* (No. arXiv:2507.19792). arXiv. https://doi.org/10.48550/arXiv.2507.19792

Drolsbach, C. P., Solovev, K., & Pröllochs, N. (2024). Community notes increase trust in fact-checking on social media. *PNAS Nexus*, *3*(7), pgae217. https://doi.org/10.1093/pnasnexus/pgae217

Fay, N., Ransom, K., Walker, B., Howe, P., Perfors, A., & Kashima, Y. (2025). Truth Over Falsehood: Experimental Evidence on What Persuades and Spreads. *Journal of Personality and Social Psychology*. https://doi.org/10.31234/osf.io/9ezx8

Fessler, D. M. T., Pisor, A. C., & Navarrete, C. D. (2014). Negatively-Biased Credulity and the Cultural Evolution of Beliefs. *PLOS ONE*, *9*(4), e95167. https://doi.org/10.1371/journal.pone.0095167

Franzen, A., & Mader, S. (2023). The power of social influence: A replication and extension of the Asch experiment. *PLOS ONE*, *18*(11), e0294325. https://doi.org/10.1371/journal.pone.0294325

Germano, F., Gómez, V., & Sobbrio, F. (2022). *Ranking for Engagement: How Social Media Algorithms Fuel Misinformation and Polarization* (SSRN Scholarly Paper No. 4257210). Social Science Research Network. https://doi.org/10.2139/ssrn.4257210

Gong, W., Lim, E.-P., & Zhu, F. (2015). Characterizing Silent Users in Social Media Communities. *Proceedings of the International AAAI Conference on Web and Social Media*, *9*(1), 140–149. https://doi.org/10.1609/icwsm.v9i1.14582

Haidt, J. (2022). Why the Past 10 Years of American Life Have Been Uniquely Stupid. *The Atlantic*.

Haidt, J. (2025). *Life After Babel: Adapting to a World We May Never Share Again*. Penguin Press.

Haun, D. M., Rekers, Y., & Tomasello, M. (2012). Majority-Biased Transmission in Chimpanzees and Human Children, but Not Orangutans. *Current Biology : CB*, *22*(8), 727–731.

Howe, P. D. L., Perfors, A., Ransom, K. J., Walker, B., Fay, N., Kashima, Y., Saletta, M., & Dong, S. (2024). Self-certification: A novel method for increasing sharing discernment on social media. *PLOS ONE*, *19*(6), e0303025. https://doi.org/10.1371/journal.pone.0303025

Jannach, D., Zanker, M., Felfernig, A., & Friedrich, G. (2010). *Recommender Systems: An Introduction*. Cambridge University Press. https://doi.org/10.1017/CBO9780511763113

Jucks, R., & Thon, F. M. (2017). Better to have many opinions than one from an expert?

Social validation by one trustworthy source versus the masses in online health forums. *Computers in Human Behavior*, *70*, 375–381. https://doi.org/10.1016/j.chb.2017.01.019

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

Kendal, R. L., Boogert, N. J., Rendell, L., Laland, K. N., Webster, M., & Jones, P. L. (2018). Social Learning Strategies: Bridge-Building between Fields. *Trends in Cognitive Sciences*, *22*(7), 651–665. https://doi.org/10.1016/j.tics.2018.04.003

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*, 1–26. https://doi.org/10.18637/jss.v082.i13

Lewandowsky, S., Cook, J., Fay, N., & Gignac, G. E. (2019). Science by social media: Attitudes towards climate change are mediated by perceived social consensus. *Memory & Cognition*, *47*(8), 1445–1456. https://doi.org/10.3758/s13421-019-00948-y

Luo, M., Hancock, J. T., & Markowitz, D. M. (2022). Credibility Perceptions and Detection Accuracy of Fake News Headlines on Social Media: Effects of Truth-Bias and Endorsement Cues. *Communication Research*, *49*(2), 171–195. https://doi.org/10.1177/0093650220921321

McLoughlin, K. L., & Brady, W. J. (2024). Human-algorithm interactions help explain the spread of misinformation. *Current Opinion in Psychology*, *56*, 101770. https://doi.org/10.1016/j.copsyc.2023.101770

Mess, S. A., Bharti, G., Newcott, B., Chaffin, A. E., Van Natta, B. W., Momeni, R., & Swanson, S. (2019). To Post or Not to Post: Plastic Surgery Practice Marketing, Websites, and

Social Media? *Plastic and Reconstructive Surgery Global Open*, *7*(7), e2331. https://doi.org/10.1097/GOX.0000000000002331

Narayan, A. (2023). *Understanding Social Media Recommendation Algorithms*. Knight First Amendment Institute. http://knightcolumbia.org/content/understanding-social-media-recommendation-algorithms

Pennycook, G., & Rand, D. G. (2022). Accuracy prompts are a replicable and generalizable approach for reducing the spread of misinformation. *Nature Communications*, *13*(1), 2333. https://doi.org/10.1038/s41467-022-30073-5

Pfänder, J., & Altay, S. (2025). Spotting false news and doubting true news: A systematic review and meta-analysis of news judgements. *Nature Human Behaviour*, 1–12. https://doi.org/10.1038/s41562-024-02086-1

Prike, T., Butler, L. H., & Ecker, U. K. H. (2024). Source-credibility information and social norms improve truth discernment and reduce engagement with misinformation online. *Scientific Reports*, *14*(1), 6900. https://doi.org/10.1038/s41598-024-57560-7

R Core Team. (2013). *R: A Language and Environment for Statistical Computing* [Computer software]. R Foundation for Statistical Computing. http://www.R-project.org/

Rathje, S., Roozenbeek, J., Van Bavel, J. J., & van der Linden, S. (2023). Accuracy and social motivations shape judgements of (mis)information. *Nature Human Behaviour*, *7*(6), 892–903. https://doi.org/10.1038/s41562-023-01540-w

Ray, G., McDermott, C. D., & Nicho, M. (2024). Cyberbullying on Social Media: Definitions, Prevalence, and Impact Challenges. *Journal of Cybersecurity*, *10*(1), tyae026. https://doi.org/10.1093/cybsec/tyae026

Rossi, W. S., Polderman, J. W., & Frasca, P. (2022). The Closed Loop Between Opinion

Formation and Personalized Recommendations. *IEEE Transactions on Control of Network Systems*, *9*(3), 1092–1103. https://doi.org/10.1109/TCNS.2021.3105616

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

Scott, C. F., Bay-Cheng, L. Y., Prince, M. A., Nochajski, T. H., & Collins, R. L. (2017). Time spent online: Latent profile analyses of emerging adults' social media use. *Computers in Human Behavior*, *75*, 311–319. https://doi.org/10.1016/j.chb.2017.05.026

Shin, I. (2022). *Twitter and Endorsed (Fake) News: The Influence of Endorsement by Strong Ties, Celebrities, and a User Majority on Credibility of Fake News During the COVID-19 Pandemic*.

Shin, I., Wang, L., & Lu, Y.-T. (2022). Twitter and endorsed (fake) news: The influence of endorsement by strong ties, celebrities, and a user majority on credibility of fake news during the COVID-19 pandemic. *International Journal of Communication*, *16*, 2573–2595.

Slaughter, I., Peytavin, A., Ugander, J., & Saveski, M. (2025). Community notes reduce engagement with and diffusion of false information online. *Proceedings of the National Academy of Sciences*, *122*(38), e2503413122. https://doi.org/10.1073/pnas.2503413122

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic Vigilance. *Mind & Language*, *25*(4), 359–393. https://doi.org/10.1111/j.1468-0017.2010.01394.x

Tenenboim, O. (2022). Comments, Shares, or Likes: What Makes News Posts Engaging in Different Ways. *Social Media + Society*, *8*(4), 20563051221130282.

https://doi.org/10.1177/20563051221130282

Traberg, C. S., Harjani, T., Roozenbeek, J., & van der Linden, S. (2024). The persuasive effects of social cues and source effects on misinformation susceptibility. *Scientific Reports*, *14*(1), Article 1. https://doi.org/10.1038/s41598-024-54030-y

Vlasceanu, M., & Coman, A. (2022). The impact of social norms on health-related belief update. *Applied Psychology: Health and Well-Being*, *14*(2), 453–464. https://doi.org/10.1111/aphw.12313

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

Wirtschafter, V., & Majumder, S. (2023). Future Challenges for Online, Crowdsourced Content Moderation: Evidence from Twitter's Community Notes. *Journal of Online Trust and Safety*, *2*(1). https://doi.org/10.54501/jots.v2i1.139