Filtering the Evidence: How Communicative Goals Shape the Balance of Opinion About Health Claims on Social Media

Benjamin Simmonds (benjamin.simmonds@adelaide.edu.au)[1], Rachel G. Stephens[1] & Keith J. Ransom[2]

[1]School of Psychology, College of Education, Behavioural and Social Sciences, Adelaide University

[2]School of Computer Science and Information Technology, College of Engineering and Information Technology, Adelaide University

Abstract

Increasingly, people encounter health-related claims made by others on social media. When faced with conflicting information, an important question is how people decide which information to share. Selective sharing (e.g., passing on only posts that support a particular claim) may distort others' perceptions of the balance of opinion. Across two experiments, we examined how communicative goals shape information sharing by asking participants to choose sets of posts to share with different goals in mind. In Experiment 1, we found distinct sharing behaviours across goals, and that when given no explicit goal, people tended to share as if advocating for their preferred stance rather than conveying the balance of opinion in a more nuanced way. Experiment 2 replicated these results even when the balance of opinion opposed people's own beliefs. This research highlights how biases in information sampling may systematically distort social learning.

Keywords: social media; health beliefs; sharing behaviour; information sampling.

Introduction

When evaluating the truthfulness of a disputed health claim, such as "*salt lamps improve sleep*", it may be difficult or even impossible to gather direct evidence to inform ourselves. In these instances, we can instead turn to those around us in the hope that they can provide relevant information, leveraging a process called *social learning* (Boyd et al., 2011). Social media platforms such as X, Facebook, Instagram, and Reddit, facilitate social learning by increasing the ease with which people can gather and share health information. People take considerable advantage of this connectedness by using social media as an information base to guide their health behaviours (National Cancer Institute, 2024).

But social media has the potential to distort social learning by changing the way that information is *mediated* – that is, selectively sampled and passed on by a person who transfers information from its source to a recipient. Traditionally, occupying this role at scale was the preserve of large institutions. For example, news media outlets have traditionally mediated the communication between health organisations and the public (Metzger, 2007). Today, however, social media platforms allow users to share, forward, or repost content that can quickly spread across a potentially global network. As social media becomes a critical resource for health information (Lim et al., 2022), there is growing potential for mediators' sampling biases to shape recipients' inferences. Understanding how people choose what to share when acting as information intermediaries is therefore increasingly important.

A particular influence that the mediator can have is in shaping the recipient's perception of the balance of opinion concerning a health claim. Others' opinion about whether a claim is true may represent valuable information when evaluating the claim's veracity. For example, a broad consensus in favour of a claim may be viewed as strong evidence, since agreement between multiple unbiased and independent sources would be unlikely if their shared conclusion were false (Huemer, 2005). Prior research suggests that people recognise the value in this type of social information: information recipients often revise their beliefs in proportion to the balance of opinion observed (Ransom et al., 2021), and tend to agree more with claims supported by a strong scientific consensus (Chinn et al., 2018).

A key question is how mediators decide what balance of supporting and dissenting information to share. A mediator acting without reference to their own beliefs, for example, might choose to impartially *reflect* the balance of others' opinions by sharing supporting and dissenting posts in proportion to their prevalence online. But mediators may also sample posts in ways that do not preserve the observed balance of opinion, potentially distorting others' impression of the prevailing consensus. One such strategy is that of a simplified Bayesian-like mediator who incorporates evidence inferred on the basis of others' opinions with their own "first-hand" evidence (that is, evidence that was not socially inferred), effectively sharing supporting and dissenting opinions to reflect the strength of their posterior belief (i.e., their own *certainty* in the claim). For example, a mediator who views the balance of opinion and judges a claim to be 75% likely to be true and 25% likely to be false may express this belief by sharing supporting information 75% of the time and opposing information 25% of the time. This type of *probability matching* is a strategy observed in other decision-making contexts (Saldana et al., 2022; Stanovich, 2003).

Another strategy a mediator may use is to share supporting and opposing information with equal frequency. Traditionally, this approach has been used by professional journalists to adhere to norms of objectivity and fairness (Bennett, 1996), and may reflect a well-intentioned aim both to provide an *evenly balanced* overview of a contested issue, and to be seen as impartial and thereby reduce reflexive dismissal of the information by an audience sensitive to bias. However, if the actual balance of opinion

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

concerning a claim is *not* evenly balanced, sharing information in this way can create a "false balance" (Boykoff & Boykoff, 2004), elevating fringe or contrarian perspectives, giving them disproportionate influence.

Alternatively, a mediator may share information that disproportionately supports a particular stance or perspective. A mediator who intends to *advocate* for a particular stance may use this strategy to shift recipients' beliefs in favour of that stance (cf. pedagogical sampling, Shafto et al., 2014). This parallels how social media algorithms preferentially present belief-consistent content, contributing to the formation of 'filter bubbles' (Bozdag, 2013). By similarly lessening exposure to counter-attitudinal information, the mediator constrains the recipient's view, potentially giving the impression that a consensus exists where it does not.

Despite the potential for mediators to shape perceptions of the balance of opinion, there is limited research examining how people behave when sharing others' opinions, or which, if any, of these strategies they might employ. One exception is a study by Perfors et al. (2018), who found that information mediators tend to disproportionately share dissenting opinions from an almost unanimous consensus. The authors interpreted this as reflecting the mediator's assumptions about the recipient - namely, that the recipient would consider the mediator to be biased if they shared only confirmatory information. However, the participants were instructed to imagine themselves as journalists, which may have implied that the goal was to come across as unbiased. The current research builds on this earlier work by systematically manipulating the participant's communicative goal, with the aim of examining the impact on their sharing decisions.

Experiment 1

The aim of our pre-registered experiment[1] was to examine how mediators choose to convey the balance of opinions about health-related claims under four different communicative goals. A control condition, in which no explicit goal was assigned, was also included to examine which goal participants would naturally adopt in an online context.[2]

Method

Participants We recruited 205 participants from Amazon's Mechanical Turk (MTurk) in October 2025. 52 additional participants were recruited in November 2025 to add the CERTAINTY condition[3]. All participants were at least 18 years of age and were pre-screened for English proficiency. They received $3USD for approximately 10-15 minutes of participation. 33 participants were excluded from data analysis according to a pre-registered attention check. The remaining 224 participants were randomly assigned to one of five conditions: BALANCE (N = 42), ADVOCATE (N = 46), REFLECT (N = 47), CERTAINTY (N = 39), and CONTROL (N = 50).

Among this sample, 114 (50.89%) identified as male, 107 (47.77%) identified as female, and 3 (1.34%) identified as neither. Ages ranged from 21 to 79 years old (M = 43.87, Med = 42.50, SD = 10.96). The majority of participants resided in the United States (80.36%).

Materials – Health Claims Participants viewed five health claims, one per trial (e.g., "*Medicinal cannabis reduces symptoms of depression*"). Claims were taken from previous work by Simmonds et al. (2023, 2024), who originally compiled them from online fact-checking websites such as Snopes.com. Two claims were true and three were false[4].

Materials - Social Media Posts For each claim, we created 16 short social media posts that either supported or opposed it. Each post was a simple statement of support or opposition without any accompanying argument or evidence and was framed as a social media post (see Figure 1b). We phrased each slightly differently to give the impression that they had been written by unique sources. Among each set of 16, 12 always aligned with the ground truth (i.e., supported true claims and opposed false claims), and four always aligned against the ground truth (i.e., opposed true claims and supported false claims). Posts were generated either manually or using GPT-5.

Materials - Sources Each post's source was a unique mock social media user. Participants could view the source's display name, avatar, and handle, all of which were adapted from Simmonds et al. (2023, 2024).

Design We manipulated the participant's communicative goal between-subjects across five levels: REFLECT, BALANCE, ADVOCATE, CERTAINTY, and CONTROL. In the REFLECT condition, participants were asked to select four posts that, if someone saw only those four, would reflect the underlying proportion of posts supporting and opposing the claim. In the BALANCE condition, participants were asked to select four posts that, if someone saw only those four, would display a balance of opinions and perspectives. In the ADVOCATE condition, participants were asked to select four posts that, if someone saw only those four, would convince them to adopt the participant's view of the claim. In the CERTAINTY condition, participants were asked to select four posts that, if someone saw only those four, would best reflect how much the participant agreed with the claim after reading all of the posts. Finally, in the CONTROL condition, participants were asked to select four posts that, taken as a set, they would want to share online. The CONTROL condition served as a point of comparison to the other conditions to examine which communicative goal participants may have in mind when not assigned one.

Procedure Participants were informed that they would be asked to rate their beliefs in a series of health claims before

[1]https://aspredicted.org/s7n7-h67t.pdf

[2]This research received ethics approval by the Human Research Ethics Subcommittee, School of Psychology, University of Adelaide (Ref. 22/75).

[3]https://aspredicted.org/da9mj5.pdf

[4]see https://osf.io/9jfk8/overview?view_only=b0f3f2cc75a0443cb698882220e32aa1 for the full set of claims.

a) b)

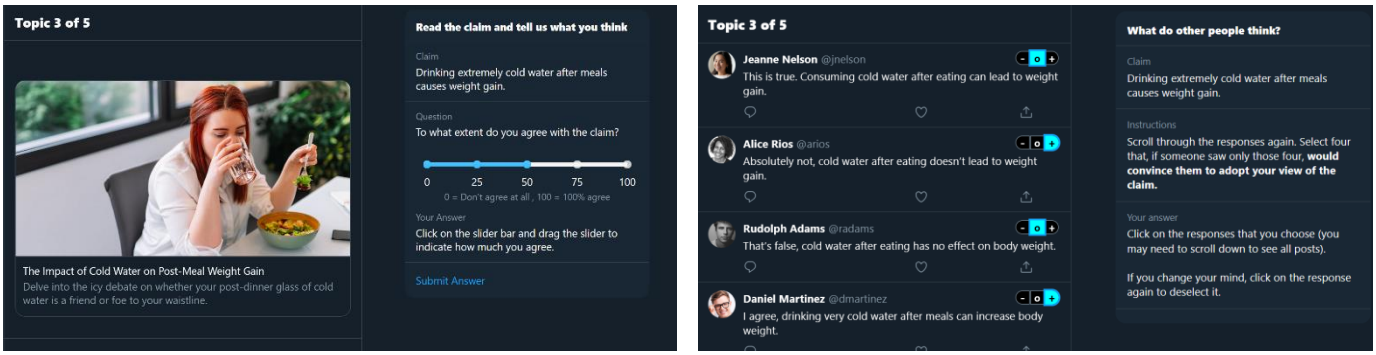

Figure 1: a) Screenshot of the experimental platform when participants provided their prior belief rating. b) Screenshot of the experimental platform when participants assigned to the ADVOCATE condition made their sharing selections.

and after viewing posts from various social media users. Their assigned goal was described, then they were given five validation questions about the instructions which they needed to answer correctly to progress. Participants then provided basic demographic information, including age, gender, country, race, education level, social media use, and political leaning.

The experiment consisted of five trials. In each trial, participants viewed a randomly selected health claim and provided a prior belief rating on a sliding scale ranging from 0 ("Don't agree at all") to 100 ("100% agree"), as shown in Figure 1a. After submitting their initial rating, participants were shown 16 randomly ordered social media posts addressing the claim. As an attention check, participants were asked to classify each post according to whether it agreed with, disagreed with, or was neutral toward the claim.

Next, participants were given the assigned goal and asked to select a subset of four posts that would meet the goal when shared with another person (see Figure 1b). This trial design meant that the four selected messages would always reflect one of five possible compositions of posts that agree or disagree with the ground truth (0:4, 1:3, 2:2, 3:1, 4:0). After confirming their selection, participants were asked to update their belief rating (their prior rating was visible), before proceeding to the next trial.

The first trial was a practice trial to allow for initial task exploration and always presented the claim that *"Salt lamps reduce anxiety, improve sleep, and increase energy*". Data from this trial were excluded from analysis. The remaining four trials were presented in random order.

Results

Preliminary Analyses Prior to examining participants' sharing selections, we were interested in the extent to which the posts affected their belief ratings. Both prior belief ratings and posterior belief ratings were standardised such that a higher value indicates greater alignment with the ground truth (i.e., the majority stance). On average, participants' ratings aligned with the ground truth prior to viewing the posts (*Median* = 75, *IQR* = 40.25), and there was a shift in its favour after viewing the posts (*Median* = 85, *IQR* = 34). A Wilcoxon signed-rank test suggested that this shift was statistically significant and had a large effect size[5] ($V = 22183$, $p < .001$, $r = .53$). Therefore, participants tended to shift their beliefs in accordance with the ground truth / majority stance.

Sharing Behaviour The red bars in Figure 2a illustrates the proportion of trials for which people selected a given mix of posts supporting or opposing the ground truth. In the BALANCE condition, participants most frequently shared a matching number of supporting and opposing posts (70.24%). In the REFLECT condition, participants most frequently shared three supporting posts and one opposing post (54.79%), accurately reflecting the underlying consensus. In the ADVOCATE condition, participants most frequently shared supporting posts only (82.61%), which was also the case in the CERTAINTY condition (78.21%) and the CONTROL condition (81.50%).

In order to examine the relationship between participants' sharing behaviour, their assigned goal, and their posterior beliefs, we fit the data to four binomial generalised linear models. These models sequentially added predictors of the likelihood a participant would share a supporting post over an opposing post. Table 1 displays the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) of these models, indicating that

Table 1: Comparison of AIC and BIC across four binomial general linear models predicting the likelihood of selecting a supporting post. Model in bold is preferred.

# Predictors	AIC	BIC
1 Intercept-Only	2615.19	2619.98
2 Posterior Belief Rating	2406.52	2430.51
3 Posterior Belief Rating + Condition	1994.82	2023.61
4 Posterior Belief Rating ✕ Condition	**1698.51**	**1746.49**

[5] $r \geq .5$ is a large effect, $r \geq .3$ is a moderate effect, and $r \geq .1$ is a small effect (Fritz et al., 2012).

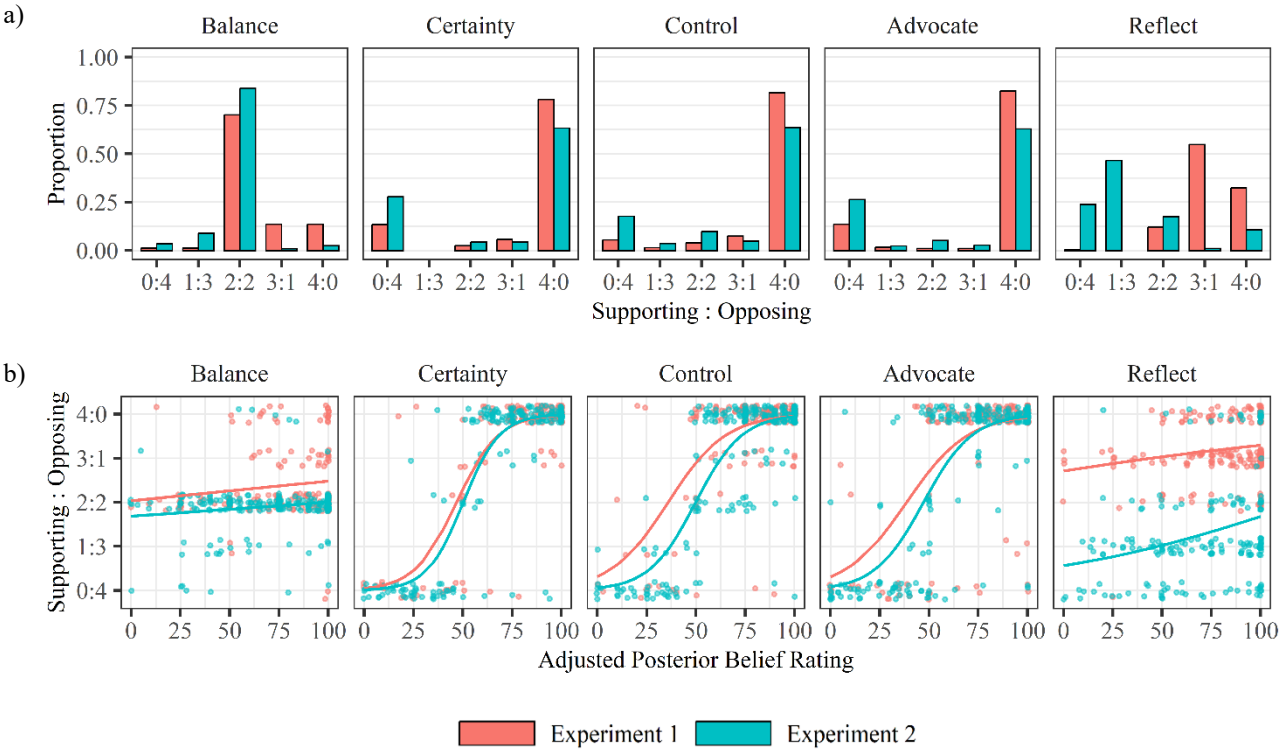

Figure 2: a) Proportion at which each possible composition of posts supporting and opposing the ground truth was shared, as a function of sharing goal condition and experiment. In Experiment 1, supporting posts were in the majority and opposing posts were in the minority; in Experiment 2, this was reversed. b) The relationship between adjusted posterior belief ratings, the composition of supporting and opposing posts shared, and the assigned sharing goal condition, separated by experiment. Regression curves represent the predictions of each experiment's preferred binomial generalised model.

the best-fitting model was Model 4, which contained the following predictors: the participant's posterior belief rating (standardised such that a higher score indicated greater alignment with the ground truth / the majority stance), the communicative goal condition that the participant was assigned to, and their interaction effect.

The red curves in Figure 2b display the predictions of Model 4, as a factor of their assigned goal. There was a main effect of posterior belief rating on the participants' sharing behaviour (b = 0.07, SE = 0.01, p < .001). People were more likely to share compositions with more supporting posts when their posterior belief aligned more with the claim. Interaction effects indicated that, with the CONTROL condition as the reference level, the association between posterior belief rating and selection composition was weaker in the BALANCE condition (b = -0.06, SE = 0.01, p < .001) and in the REFLECT condition (b = -0.06, SE = 0.01, p < .001). The association was stronger in the CERTAINTY condition (b = 0.03, SE = 0.01, p = .014) but did not significantly differ in the ADVOCATE condition (b = 0.01, SE = 0.01, p = .624).

Therefore, participants tended to most frequently share a 2:2 and 3:1 composition in the BALANCE and REFLECT conditions, respectively, regardless of their belief in the claim after viewing the posts. There was no statistically significant difference between sharing behaviour in the CONTROL condition and the ADVOCATE condition. In the CERTAINTY condition, however, there was a stronger positive association between the participants' beliefs and their sharing behaviour.

Discussion

Experiment 1 examined how people select information to share when acting as mediators with different communicative goals. As expected, the balance of opinion in the samples people selected was affected by their assigned goal.

When asked to provide an even balance of opinions and perspectives, people overwhelmingly shared an equal number of supporting and opposing posts, and when asked to reflect the underlying balance of opinion (3:1), people generally did so accurately. Both behaviours were consistent regardless of the participant's belief about the claim, demonstrating a clear recognition of how to employ these strategies. When asked to advocate for their preferred stance, people overwhelmingly did so through sharing only posts that expressed the stance that aligned with their own. Similar behaviour was observed both when participants were asked to reflect their degree of certainty in the claim, and, of particular interest, in the CONTROL condition, where people were asked to make selections as if they were sharing online. Indeed, participants' sharing behaviour in the CONTROL condition was not significantly different from sharing behaviour in the ADVOCATE condition. What this

suggests is that, when not assigned a communicative goal, people tend to advocate for their preferred stance.

Notably, advocacy led to the consistent exclusion of contrary minority opinions when sharing. That we observed this behaviour in the CONTROL condition raises concerns regarding how mediators may selectively share in a manner that leads to opposing views being filtered out of the conversation. In the current experiment, however, this filtering effect was not necessarily harmful as the excluded posts were largely opposed to the ground truth. What is more important to examine, however, is whether this effect persists when the minority opinion reflects the ground truth, and is contested by a large, incorrect majority. Will people still choose to exclusively share posts they believe to be true, even when those posts are in the minority, or will they be swayed by the influence of a large majority whose opinion they believe to be false? Experiment 2 sought to investigate this question.

Experiment 2

Experiment 2 was identical to Experiment 1 with one key difference: the four minority posts now supported the ground truth, and the 12 majority posts opposed the ground truth[6].

Method

Participants We recruited 277 participants via Amazon Mechanical Turk (MTurk) in January 2026. Requirements and compensation for participation were identical to Experiment 1. 31 participants were excluded for failing our pre-registered attention check, leaving 246 participants, randomly assigned to one of five conditions: BALANCE (N=50), ADVOCATE (N=52), REFLECT (N=44), CERTAINTY (N=45), and CONTROL (N=55).

Among this sample, roughly 148 (60.16%) identified as male, 95 (38.62%) identified as female, and 3 (1.22%) identified as neither. Ages ranged from 22 to 82 years old (M = 44.19, Med = 42, SD = 11.75). The majority of participants resided in the United States (84.15%).

Results

Preliminary Analyses Per Experiment 1, we examined how participants revised their beliefs after viewing the posts. As in the previous experiment, participants' prior ratings tended to align with the ground truth ($Median$ = 75, IQR = 45), but this time there was no change in the median posterior rating ($Median$ = 75, IQR = 39). Despite this, a Wilcoxon signed-rank test suggested that there was a statistically significant decrease in belief ratings after participants viewed the posts (V = 164586, p < .001), with a moderate effect size (r = .32). Therefore, much like in Experiment 1, participants shifted their beliefs towards the majority, but unlike Experiment 1, this led to decreased belief in the ground truth.

Sharing Behaviour As the blue bars in Figure 2a indicate, participants assigned to the BALANCE condition most frequently shared an equal number of supporting and opposing posts (84.00%), and participants assigned to the REFLECT condition most often shared a composition that accurately reflected the 1:3 consensus (46.59%). In the ADVOCATE, CERTAINTY, and CONTROL conditions, participants most often shared four supporting posts (4:0), despite these posts now being in the minority.

We fit the same four binomial general linear models from Experiment 1 to the data. AIC and BIC values are shown in Table 2, which indicated a preference for Model 4.

The blue curves in Figure 2b visualise the predictions of this model as a factor of the participants' assigned goal. There was a significant main effect of posterior belief rating (b = 0.08, SE = 0.01, p < .001), with greater belief in the ground truth leading to a greater likelihood of sharing posts that support the ground truth. With the CONTROL condition as the baseline, interaction effects indicated that the relationship between posterior beliefs and sharing behaviour was weaker in the BALANCE condition (b = -0.08, SE = 0.01, p < .001) and in the REFLECT condition (b = -0.07, SE = 0.01, p < .001). The relationship was significantly stronger in the CERTAINTY condition (b = 0.02, SE = 0.01, p = .017) but was not significantly different in the ADVOCATE condition (b = -0.01, SE = 0.01, p = .620).

Exploratory Analyses We compared sharing behaviour between Experiments 1 and 2 by fitting a binomial generalised model to the pooled data. We excluded data from the REFLECT and BALANCE conditions from this analysis, as we were interested in assessing whether the stance of the majority had a significant influence on the tendency to exclusively share a single opinion, as seen in the ADVOCATE, CERTAINTY, and CONTROL conditions. The model included experiment type as a predictor, while also controlling for adjusted posterior belief rating and assigned goal. There was a significant main effect of experiment type on sharing behaviour, such that participants were more likely to share opposing posts in Experiment 2 compared to Experiment 1 (b = -1.79, SE = 0.53, p < .001), when controlling for posterior beliefs.

Overall, the results of Experiment 2 resemble those of Experiment 1: participants generally selected a balanced composition in the BALANCE condition, tended to accurately reflect the consensus in the REFLECT condition, and advocated for their preferred stance in the ADVOCATE, CERTAINTY, and CONTROL conditions. However, we also observed some influence of the incorrect majority in Experiment 2, which affected both participants' beliefs and their advocacy-based choices when sharing.

Table 2: Comparison of AIC and BIC across four binomial general linear models predicting the likelihood of selecting a supporting post. Model in bold is preferred.

# Predictors	AIC	BIC
1 Intercept-Only	4080.98	4085.87
2 Posterior Belief Rating	3699.27	3723.73
3 Posterior Belief Rating + Condition	2688.79	2718.14
4 Posterior Belief Rating × Condition	**2145.51**	**2194.42**

[6]Pre-registration: https://aspredicted.org/a6xf62.pdf

Discussion

Experiment 2 examined how people select information to share according to different communicative goals, but asked participants to do so in the face of a majority opinion that opposed the ground truth. We found that people demonstrated sharing behaviour much like that observed in Experiment 1, with some notable differences.

As in Experiment 1, when asked to provide a balance of opinions, or when asked to reflect the underlying consensus, participants generally did so in an accurate manner regardless of their beliefs. Furthermore, when asked to advocate for their preferred stance, convey their certainty in the claim, or share as if on social media, participants continued to prioritise the sharing of a subset of posts that exclusively aligned with their stance.

A notable finding was that participants were affected by the presence of an opposing majority; participants' belief ratings significantly shifted toward the incorrect majority, and they were, on average, less likely to share posts supporting the ground truth. Interestingly, while the presence of the incorrect majority led to more moderate belief towards the claims, participants' sharing behaviour in the CERTAINTY and CONTROL conditions did not reflect this degree of moderation. Instead, they continued to predominantly share only supporting posts (4:0) or opposing posts (0:4), depending on their stance on the claim.

General Discussion

Across two experiments, we examined how communicative goals shaped the information people choose to share when acting as mediators in a social media-like environment. There were several important findings. First, we found that in the absence of an assigned goal (CONTROL condition), people appeared to *advocate* for their preferred stance, rather than providing an evenly balanced sample or reflecting the proportion of opinions observed. This suggests that Perfors et al.'s (2018) findings – that people tend to disproportionately include dissenting posts – may have been a byproduct of the journalism cover story which encouraged a more balanced selection.

Furthermore, when advocating, people tended to do so by exclusively sharing pro-attitudinal content. In both experiments, this was relatively benign, as people largely advocated for the ground truth by filtering those that opposed it. However, when mediators are uninformed or misinformed, the filtering of contrary opinions on social media may facilitate the trapping of other users into harmful 'filter bubbles' (Bozdag, 2013).

These findings were generally consistent regardless of whether the participant's preferred stance was supported or opposed by a majority. With that said, the majority still held some influence over the participants' beliefs, ultimately affecting which stance was advocated for. In some instances, this led to participants advocating against the ground truth. Nevertheless, people's sharing behaviour did not show the same high sensitivity to the balance of others' opinions that has been observed in prior work on consensus-based belief revision (e.g., Ransom et al., 2021; Simmonds et al., 2023, 2024). It is important to acknowledge some potential limitations associated with this research. First of all, both experiments examined sharing behaviour in response to four claims that tended to elicit relatively strong prior beliefs. It is therefore unclear whether people would still feel inclined to advocate for their preferred stance when holding weaker priors. Future research should replicate this research with claims that elicit a wider range of prior beliefs.

A further potential limitation concerns the CERTAINTY condition in Experiment 2, where we unexpectedly found that people acted as if they were advocating for their preferred stance. This was less surprising in Experiment 1, where posterior beliefs were strongly in favour of the ground truth, but more so in Experiment 2, where posterior beliefs were more moderate. If participants were probability matching, we would expect them to share a mixture of supporting and opposing posts (Saldana et al., 2022; Stanovich, 2003). Participants therefore may have misunderstood their communicative goal - potentially reflecting an issue with the instructions – or this may imply a genuine tendency to overrepresent their degree of certainty when sharing.

This research highlights how social media users can mediate health information in ways that can be both beneficial and misleading. When aiming to impartially reflect the balance of opinion, or when advocating for the ground truth, this can be broadly beneficial in informing people's health beliefs. However, mediators can also distort the perceived balance of opinion in a way that, unintentionally or not, can reinforce health-related misconceptions. Because sharing behaviour does not necessarily reflect belief, interventions solely aimed at shifting beliefs may be insufficient in curbing this filtering effect. Instead, effective interventions should target sharing behaviour as a separate variable, encouraging users to represent the broader information environment with relative accuracy.

Acknowledgements

This research was supported by funding from The University of Adelaide. The first author was supported by a Research Training Program Scholarship.

References

Bennett, W. L. (1996). An introduction to journalism norms and representations of politics. *Political Communication*, *13*(4), 373–384. https://doi.org/10.1080/10584609.1996.9963126

Boyd, R., Richerson, P. J., & Henrich, J. (2011). The cultural niche: Why social learning is essential for human adaptation. *Proceedings of the National Academy of Sciences*, *108*(supplement_2), 10918–10925. https://doi.org/10.1073/pnas.1100290108

Boykoff, M. T., & Boykoff, J. M. (2004). Balance as bias: Global warming and the US prestige press. *Global Environmental Change*, *14*(2), 125–136. https://doi.org/10.1016/j.gloenvcha.2003.10.001

Bozdag, E. (2013). Bias in algorithmic filtering and personalization. *Ethics and Information Technology*, *15*(3), 209–227. https://doi.org/10.1007/s10676-013-9321-6

Chinn, S., Lane, D. S., & Hart, P. S. (2018). In consensus we trust? Persuasive effects of scientific consensus communication. *Public Understanding of Science*, *27*(7), 807–823. https://doi.org/10.1177/0963662518791094

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, *141*(1), 2–18. https://doi.org/10.1037/a0024338

Huemer, M. (2005). Is Critical Thinking Epistemically Responsible? *Metaphilosophy*, *36*(4), 522–531. https://doi.org/10.1111/j.1467-9973.2005.00388.x

Lim, M. S. C., Molenaar, A., Brennan, L., Reid, M., & McCaffrey, T. (2022). Young Adults' Use of Different Social Media Platforms for Health Information: Insights From Web-Based Conversations. *Journal of Medical Internet Research*, *24*(1), e23656. https://doi.org/10.2196/23656

Metzger, M. J. (2007). Making sense of credibility on the Web: Models for evaluating online information and recommendations for future research. *Journal of the American Society for Information Science and Technology*, *58*(13), 2078–2091. https://doi.org/10.1002/asi.20672

National Cancer Institute. (2024). *Health Information National Trends Survey 7* [Dataset]. https://hints.cancer.gov/

Perfors, A., Navarro, D. J., & Shafto, P. (2018). Stronger evidence isn't always better: A role for social inference in evidence selection and interpretation. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *40*(0). https://escholarship.org/uc/item/9f69q5wm

Ransom, K. J., Perfors, A., & Stephens, R. (2021). Social meta-inference and the evidentiary value of consensus. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *43*(43). https://escholarship.org/uc/item/7x16q1r4

Saldana, C., Claidière, N., Fagot, J., & Smith, K. (2022). Probability matching is not the default decision making strategy in human and non-human primates. *Scientific Reports*, *12*, 13092. https://doi.org/10.1038/s41598-022-16983-w

Shafto, P., Goodman, N. D., & Griffiths, T. L. (2014). A rational account of pedagogical reasoning: Teaching by, and learning from, examples. *Cognitive Psychology*, *71*, 55–89. https://doi.org/10.1016/j.cogpsych.2013.12.004

Simmonds, B. P., Ransom, K. J., & Stephens, R. (2024). Navigating Health Claims on Social Media: Reasoning from Consensus Quantity and Expertise. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *46*(0). https://escholarship.org/uc/item/76j8m25k

Simmonds, B. P., Stephens, R., Searston, R. A., Asad, N., & Ransom, K. J. (2023). The Influence of Cues to Consensus Quantity and Quality on Belief in Health Claims. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *45*(45). https://escholarship.org/uc/item/73t0j4tc

Stanovich, K. E. (2003). Is probability matching smart? Associations between probabilistic choices and cognitive ability. *Memory & Cognition*, *31*(2), 243–251. https://doi.org/10.3758/bf03194383