Navigating Health Claims on Social Media: Reasoning from Consensus Quantity and Expertise

Benjamin Simmonds (benjamin.simmonds@adelaide.edu.au)
School of Psychology, Faculty of Health and Medical Sciences, The University of Adelaide

Keith J. Ransom (keith.ransom@adelaide.edu.au)
School of Psychology, Faculty of Health and Medical Sciences, The University of Adelaide

Rachel G. Stephens (rachel.stephens@adelaide.edu.au)
School of Psychology, Faculty of Health and Medical Sciences, The University of Adelaide

Abstract

When assessing the quality of health information encountered online, reasoners may rely on the wisdom of others and the degree of consensus apparent. However, it is unclear whether reasoners weigh the opinions of others evenly or make assumptions about the amount of evidence that each has seen. We investigated this question in an online experiment where people were asked to rate their belief in a series of health claims both before and after reading responses from other users. The degree of consensus among these users and their level of expertise (non-experts vs. expert organisations) was manipulated within-subjects. While we found belief change increased monotonically with the degree of consensus for both experts and non-experts, our results indicate qualitatively different patterns of increase between the two groups. Our study suggests that people reason from consensus using nuanced assumptions about the evidence underlying other people's opinions.

Keywords: reasoning; consensus; expertise; health information; social media.

Introduction

Social media plays a highly influential role in today's information landscape, serving as an alternative to traditional media amidst declining social capital and trust in private and public institutions (Fletcher & Park, 2017; Lewandowsky, Ecker & Cook, 2017). Social media facilitates rapid public health communication, but the proliferation of health-related misinformation on social media is a sizeable threat to the health and safety of the general public (Suarez-Lledo & Alvarez-Galvez, 2021; Zhao, Hu, Zhou, Song, Wang, Zheng, Zhang & Hou, 2023). Take the following claim: "Poor sleep is linked to Alzheimer's". Given the prevalence of misinformation and wide range of expertise on social media, how might a user determine the claim's veracity if encountered online?

Lacking sufficient expertise of their own, people often look towards the judgements of others and their degree of consensus (Asch, 1956; Harkins & Petty, 1981; Ransom, Perfors & Stephens, 2021; Simmonds, Stephens, Searston, Asad & Ransom, 2023). A simple consideration of consensus could involve a counting heuristic, where the persuasiveness of a consensus increases with its size (Mercier & Morin, 2019). Alternatively, people may make more complex inferences about the underlying quantity of evidence that informed a given consensus, wherein larger groups are assumed to have access to more evidence (Harkins & Petty, 1987). It is currently unclear which reasoning strategy people typically use, as both are consistent with classic consensus effects; a positive relationship between consensus quantity (i.e., the number of people that form a consensus) and its persuasiveness.

To help disentangle these strategies, we can explore the extent to which people consider evidential overlaps within a consensus. If people conceptualise a given consensus as an accumulation of underlying evidence, the presence of evidential overlaps (i.e., multiple sources having seen the same piece of evidence) should compromise its informational value compared to an otherwise equivalent consensus where no such overlaps exist (Whalen, Griffiths & Buchsbaum, 2018). If people instead leverage consensus using a simple counting heuristic, the existence of such evidential overlaps should have no effect.

One source of evidential overlap is in consensus quantity itself. This path of reasoning assumes that each source within the consensus represents a sample of underlying evidence viewed from a finite evidence space. As consensus quantity increases, the proportion of the total evidence space viewed by the consensus increases, but the probability of evidential overlap also increases. This overlap should result in a negatively accelerating curve when graphing the persuasiveness of a consensus against its size. This type of relationship has been reported across some of the literature (e.g., Asch, 1951; Bovens & Hartmann, 2004; Shu & Carlson, 2014), but others have found no such pattern, instead reporting a linear relationship between consensus quantity and persuasion (Gerard, Wilhelmy & Conolley, 1968; Nordholm, 1975).

Another source of evidential overlap is expertise. At a simplistic level, an expert can be differentiated from a non-expert by the size of the sample they are able to draw from the total evidence space (Klein, 2008). People appear to be aware of this, perceiving a single expert opinion as being equivalent to at least 10 non-expert opinions (Hornikx, Harris and Boekema, 2018), and generally finding experts to be more persuasive than non-experts (Jucks & Thon, 2017; Maddux & Rogers, 1980; Simmonds et al., 2023). However, when accumulated within a consensus, the large

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

sample drawn by each expert should result in a greater likelihood of evidential overlap. This overlap should be especially heightened when considering opinions from expert organisations, each of which represents a consensus of its own and thus should have an even larger coverage of the total evidence space. There is some evidence to suggest that people consider this potential evidence overlap when reasoning. A study by Simmonds et al. (2023) reported that a single expert organisation repeating the same argument five times was just as persuasive as five different expert organisations providing one argument each - a finding that was not replicated amongst individual non-expert sources. Similarly, Connor Desai, Xie and Hayes (2022) found that participants were just as persuaded by a single, repeatedly-cited expert organisation than four expert organisations cited once each, unless their independence was made clear.

The current paper aims to further explore the findings of Simmonds et al. (2023), investigating whether people reason from consensus by making assumptions about the underlying evidence that each source represents; comparing non-expert versus expert organisation sources. We begin by presenting a preliminary Bayesian analysis to aid in developing the intuition behind our hypotheses. We then tested these hypotheses in an experiment examining the effect of consensus quantity and expertise on people's belief in health claims.

Bayesian Analysis and Computational Simulation

Consider a blank reasoner who must update their belief in a health claim (h) upon exposure to a consensus comprising m number of Bayesian reasoners. Each Bayesian reasoner has constructed their belief via the sampling of evidence from a shared and limited evidence space. Amongst a consensus of m number of these reasoners, $n_1, \dots, n_m$ denotes the size of each sample drawn, and $k_1, \dots, k_m$ denotes the amount of *supporting* evidence within each sample. Both $n_1, \dots, n_m$ and $k_1, \dots, k_m$ can be summed to produce the total (n) and supporting (k) amount of evidence the consensus represents, respectively:

$$n = \sum_{i=1}^{m} n_i \quad \text{and} \quad k = \sum_{i=1}^{m} k_i \qquad (1)$$

In this example, we assume that all sources support h ($k > n - k$). The process by which the reasoner updates their belief in h after exposure to this consensus should then follow Bayes' Rule:
`

$$P(h|n,k) = \frac{P(n,k|h)P(h)}{P(n,k)} \qquad (2)$$

This equation can be simplified by specifying that the prior belief distribution, $P(h)$, is a beta distribution with parameters α and β, which respectively denote the number of successes and failures. Furthermore, as we assume that the relevant evidence exists as a series of Bernoulli trials, the likelihood distribution, $P(n,k|h)$, can be re-written as a Bernoulli distribution. By substituting these components into Bayes' theorem, we get the following formula:

$$P(h|n,k) = beta(h|\alpha + k, \beta + n - k) \qquad (3)$$

Equation 3 indicates that the posterior belief distribution is equal to the prior belief distribution but with parameters updated to account for the amount of evidence the consensus represents.

The effect of consensus quantity on beliefs is reflected by m, which when increased, subsequently results in an increase in n and k, which, in turn, shifts and tightens the shape of the posterior belief distribution. Expertise, on the other hand, is reflected in n_i and k_i. As Expertise increases (i.e., non-expert to individual expert to expert organisation), n_i will increase while keeping the proportion of n_i to k_i constant, tightening the posterior belief distribution.

As consensus quantity and expertise increase, so too does the probability of evidential overlap. If the reasoner believes that multiple sources are basing their opinion on a shared piece of evidence, that evidence should only be included once when determining n and k. The relationship between m and belief change should thus form a negatively accelerating curve, one that reaches its plateau at a faster rate when expertise is high.

We applied this framework to a computational simulation[1] using R version 4.2.3. to visualize our hypotheses. 99 simulated participants sampled a prior belief rating towards hypothesis h from a unform prior distribution ($beta(\alpha = 1, \beta = 1)$). Subjects were then exposed to a consensus of one to five Bayesian reasoners, all of whom had sampled from the total evidence space and supported h with complete certainty ($n = k$). The expertise of these reasoners was manipulated across two levels: Non-Expert and Expert. Expert sources were assumed to have sampled more of the total evidence space than non-experts ($n_i^{expert} > n_i^{non-expert}$). Evidential overlap was accounted for by removing duplicate pieces of evidence sampled from each source. The total amount of unique evidence the consensus sampled was then used to update the prior belief distribution of the blank reasoner, a posterior belief rating was sampled, and belief change was calculated (posterior minus prior).

Figure 1 displays belief change as a function of consensus quantity and expertise averaged across the 99 runs. Most important to this figure is the qualitative patterns that emerge. Belief change is consistently higher upon exposure to an Expert consensus (for >0 sources) than Non-Expert consensus when keeping consensus quantity constant. Across both expertise conditions, belief change also displays a negatively accelerating curve against consensus quantity. However, the rate at which belief

[1] See OSF link: https://osf.io/hkeps/?view_only=c20f7d2290864047b2014f825bf18f20

change plateaus is greater in the Expert condition than in the Non-Expert condition.

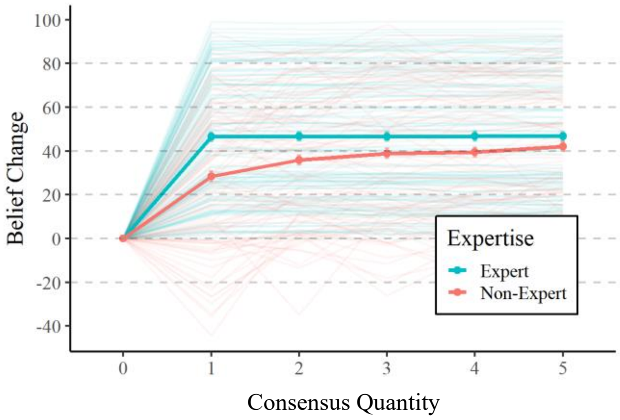

Figure 1. Simulated data run 99 times. Displays belief change as a function of consensus quantity across Non-Expert and Expert conditions. Bolded lines display mean values. Thin lines display each run.

On the basis of this analysis, we first hypothesise that in an experiment, people's beliefs will shift in the direction of the consensus position (with greater consensus quantity leading to a larger shift), consistent with prior findings (e.g., Asch, 1956; Ransom, Perfors & Stephens, 2021). We next hypothesise that Expert organisations will lead to greater belief change than Non-Expert individuals when holding consensus quantity constant. We posit that this will arise because expert organisations are assumed to have access to a greater quantity of underlying evidence to inform their opinion (Hornikx, Harris & Boekima, 2018). We also hypothesize that there will be a monotonically increasing relationship between consensus quantity and the degree of belief change. However, the difference in mean belief revision on trials with m sources compared to trials with $m-1$ sources will decrease as m increases, due to perceived evidential overlap. We further hypothesize that this will decrease at a greater rate as m increases in Expert trials than Non-Expert trials, due to the perception of larger evidential overlaps (e.g., Connor Desai, Xie & Hayes, 2022).

Method

The current experiment aimed to explore how belief in health claims changes after exposure to a consensus of varying levels of consensus quantity and expertise. In a mock social media platform, participants were shown 12 health claims across 12 trials, with one claim per trial. In each trial, participants rated their belief in the claim before being exposed to five responses from five different users. These responses either argued in favour of or against the veracity of the claim (called Target Responses) or were related but took no clear stance on its veracity (called Filler Responses). The ratio of Target to Filler Responses present, as well as the expertise of the users providing the responses were manipulated across trials (Non-Expert individuals vs Expert Organisations). Individual non-experts were compared to expert organisations to maximize the potential difference in underlying evidence coverage. The method was pre-registered on AsPredicted before data collection (https://aspredicted.org/F12_7ZG).

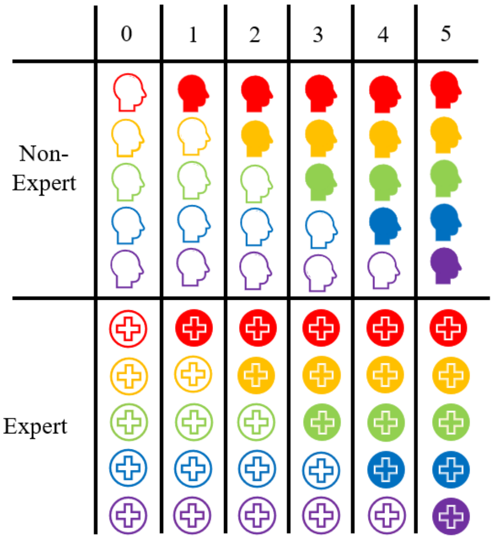

Figure 2. Experimental design. In the icons shown, each colour is a unique user; no colour fill = Filler Responses, colour fill = Target Responses. Heads = Non-Experts, crosses = Expert organisations.

Design

Trials were organised into a 6x2 factorial design (see Figure 2). Consensus quantity was manipulated by changing the ratio of Target to Filler Responses presented within-subjects across six levels: 0:5, 1:4, 2:3, 3:2, 4:1, and 5:0. The total number of responses in any given trial was always five to ensure consistency in reading time across trials. All-Filler trials (0:5) served as control conditions. The expertise of those authoring the responses was manipulated within-subjects across two levels: Non-Expert and Expert. In Expert trials, all responses were authored by health organisations. In Non-Expert trials, all responses were authored by ordinary individual users. Source diversity was controlled such that each response came from a unique source. Target Responses always argued in favour of the "ground truth", determined by the existing scientific consensus (i.e., supported true claims, argued against false claims). Argument diversity amongst the Target Responses was controlled such that, for each claim, responses would convey the same argument but slightly reworded.

Participants

139 participants were recruited on Amazon Mechanical Turk (MTurk) in January 2024. Concerns of data quality from MTurk have been raised in recent years (e.g., Ahler, Roush & Sood, 2021; Chmielewski & Kucker, 2019; Kennedy, Clifford, Burleigh, Waggoner, Jewell & Winter, 2020). To alleviate these concerns, participants were pre-screened for English-proficiency and response quality via the use of qualification tests. Participants also had to be at least 18 years of age, and each was compensated $3USD. Analysis of the pre-registered attention check resulted in the exclusion of 40 participants who scored lower than 80%,

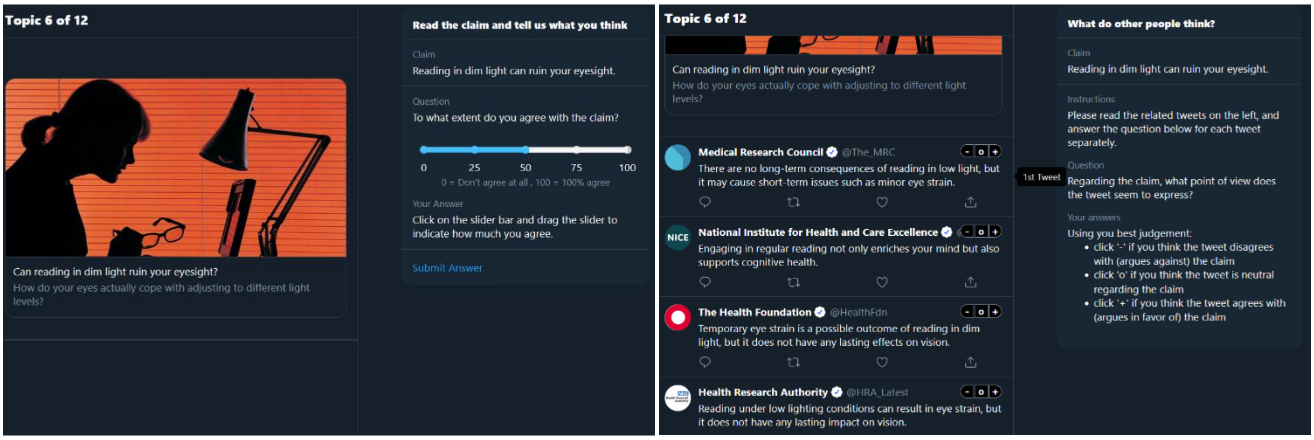

Figure 3. Screenshots of the mock social media platform. Left panel is the prior belief rating screen. Right panel is the attention check screen.

leaving a total sample size of 99. This sample size was deemed sufficient as it aligns with past studies examining online reasoning (e.g., Alister, Perfors & Ransom, 2022; Jucks & Thon, 2017; Simmonds et al., 2023). The age range of the sample was between 20 and 77 years of age (M = 40.52, *Med* = 39, SD = 10.70). The majority of participants were white, (71.72%) male, (58.59%) from the US (86.87%), had attained a high school education or above (98.99%), self-identified as left-leaning (52.51%), and used social media daily (74.75%).

Materials

Claims and Responses Participants responded to 12 claims[1] covering various health topics (e.g., "Brown sugar is healthier than white sugar."). The final set of claims used in the experiment was derived from the results of a pilot test, where claims that evoked low polarisation from respondents were selected for inclusion. These claims were primarily based on topics covered by popular fact-checking websites (e.g., Snopes), and common health myths. Half of the 12 claims had scientific backing, while the remaining half did not. This design choice was made with the aim of reflecting the variable levels of accuracy that characterise real online health claims.

Target Responses[1] were generated by writing an argument that aligned with the ground truth in response to each claim (e.g., "Regardless of its reduced processing, brown sugar is no healthier than white sugar."). ChatGPT was then used to reword this argument five times as if they were coming from five different users. Filler Responses[1] were generated by hand with further assistance from ChatGPT. These responses concerned a related health topic and shared certain keywords with the claim but did not argue for or against it (e.g., "Brushing, flossing, and limiting sugary treats are the best ways to prevent tooth decay.").

Response Sources Responses were authored by either non-experts or experts[1]. Non-expert names and avatars were fictional and randomly selected for each trial. Avatars were either AI-generated or collected online. Names were generated using random name generators. Expert groups were indicated using a verification tick and consisted of real health organisations (e.g., Australian Medical Association). We specifically chose organisations that were less well known, to minimize a potential familiarity confound. The organisations' logos were used for avatars.

Procedure

Participants were given a brief description of the study, provided informed consent, and shown task instructions. Instructions stated that for each of the 12 trials they would be shown a health claim and five responses from users of various backgrounds. It was emphasized that users with a verification tick were credible sources of subject-matter expertise. This clarification attempted to minimize any confusion surrounding our use of the verification system. Participants then completed four validation questions that required correct answers to allow further progress. They also provided basic demographic information.

Participants completed 12 trials. For each trial, a randomly selected health claim appeared above a 100-point sliding scale (see left panel of Figure 3). Participants were asked to rate the extent to which they agreed with the claim using the scale, which ranged from 0 (strongly disagree) to 100 (strongly agree). After submitting their initial belief rating, participants were shown five randomly ordered responses. To encourage participants to carefully attend to these responses, they were barred from progressing until they had labelled each in terms of its stance towards the health claim. The label options were "+", "-" and "o", which represented agreement with the claim, disagreement with the claim, and neutrality towards the claim, respectively (see right panel of Figure 3). The accuracy of these responses was later used as an attention check, where participants with <80% accuracy were excluded from analysis. This step also allowed us to ensure that the perceived stance of responses aligned with what was intended. After labelling each response, the labels disappeared and the health claim and sliding scale reappeared on screen, which participants used to provide an updated belief rating (their initial rating could be seen). After completing 12 trials, participants were debriefed and paid.

Results

We first examined participants' attention check accuracy. Out of 60 responses to label, participant accuracy ranged from 11.67% to 100% (M = 81.06%). Participants who scored less than 80% were excluded from further analysis, resulting in a final sample size of 99 with a mean accuracy of 93.67%.

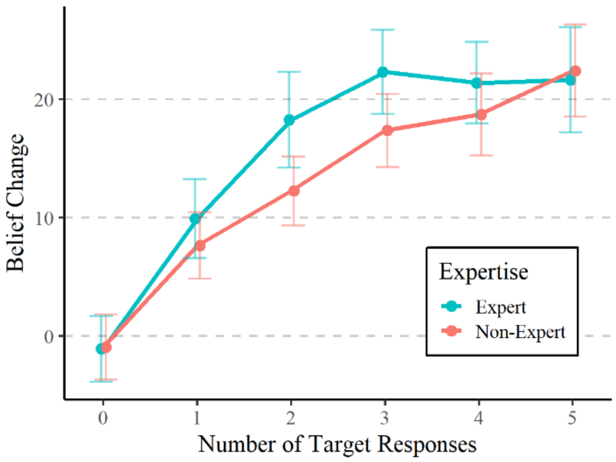

Figure 4. Belief change as a function of the number of Target Responses present across Non-Expert and Expert conditions. Points display mean values. Error bars are within-subjects 95%CI.

Figure 4 displays the relationship between belief change (posterior minus prior) and the number of Target Responses across both expertise conditions. Belief change was sign-adjusted such that a positive value reflects a shift towards the stance held by the Target Responses. In Figure 4, exposure to Target Responses tended to, on average, elicit a positive change value. Belief change also tended to increase with the number of Target Responses present. The Expert and Non-Expert curves feature some overlap, although the Expert condition evoked greater belief change at 1-4 Target Responses. Both curves appear to be negatively accelerating but the Expert curve plateaus faster than the Non-Expert curve.

Our hypotheses were statistically tested using a sequence of linear regression models. Model 1 served as a baseline model predicting belief change. Model 2 added a predictor that captured the number of Target Responses present. Model 3 added a predictor which captured the presence of expertise. Finally, Model 4 added a predictor that captured the interaction between the number of Target Responses present and their expertise. Comparison between these models was conducted using AIC and BIC (see Table 1), which showed divergent preferences: BIC most preferred Model 2, while AIC most preferred Model 3. This type of disagreement can occur as BIC more strongly penalises models that add less meaningful variables (Leppink, 2019). For the purposes of this paper, both preferred models will be outlined.

Model 2 was statistically significant ($F(2, 1185) = 108.4$, $p < .001$), with number of Target Responses being a significant predictor of belief change ($t(1185) = 14.71$, $p <$

Model	Predictors	AIC	BIC	SE
1	Prior	10386	10401	19.12
2	Prior + Target Responses	10189	**10209**	17.59
3	Prior + Target Responses + Expertise	**10184**	10210	17.55
4	Prior + Target Responses * Expertise	10186	10217	17.55

Table 1. Comparison of the four linear regression models on the basis of AIC and BIC (lower AIC and BIC indicate better fit).

.001). Model 3 was also statistically significant ($F(3,1184) = 74.73$, $p < .001$), indicating a significant effect of number of Target Responses ($t(1184) = 14.75$, $p < .001$) and expertise ($t(1184) = 2.52$, $p = .01$). Standardised regression coefficients from Model 3 indicated a greater influence of number of Target Tweets (β = 0.39) on belief change than expertise (β = 0.07). The small effect of expertise is reflected in the disagreement between BIC and AIC, indicating that the addition of expertise in Model 3 only added marginal predictive value. These results support the first hypothesis, suggesting that increasing the number of Target Responses increased belief change. They also provide tentative support for the second hypothesis, as expertise had a significant but relatively small effect on belief change.

To test the third hypothesis, we can observe Figure 4, which suggests that as the number of Target Responses increases, so too did belief change in a monotonically increasing pattern. When aggregating across expertise conditions, belief change increased as the number of Target Responses increased from 0 (M = -0.98, SD = 14.00) to 1 (M = 8.78, SD = 15.63), from 1 to 2 (M = 15.26, SD = 18.04), and from 2 to 3 (M = 19.83, SD = 17.03). After 3 Target Responses, belief change began to plateau when increasing to 4 (M = 20.04, SD = 17.56) and 5 (M = 22.01, SD = 21.15). These results support a monotonic relationship between the number of Target Responses and belief change. However, this relationship begins to display diminishing returns as consensus quantity increases, thus supporting our third hypothesis.

As Model 4 was not the most preferred model by either AIC or BIC, our findings do not fully support the fourth hypothesis. However, we note qualitative similarities between our experimental findings (Figure 4) and the predictions made in our Bayesian analysis (Figure 1), wherein the Expert condition increases before plateauing (at three Target Responses instead of one, however), while the Non-Expert condition gradually increases with consensus quantity. This indicates some support for the fourth hypothesis, but more investigation is warranted.

Discussion

We aimed to investigate whether people leverage a simple assumption about the amount of underlying evidence a consensus represents, by measuring belief change across varying levels of consensus quantity and expertise. A core finding was that consensus is a consistently persuasive cue, aligning with much of the previous literature (e.g., Harkins & Petty, 1981; Ransom, Perfors & Stephens, 2021;

Simmonds et al., 2023). Additionally, our results show that consensus quantity has a monotonically increasing relationship with belief change. This finding also aligns with literature that reports a majority to be increasingly persuasive as its size grows (e.g., Mercier & Morin, 2019).

However, we note a pattern of diminishing returns in persuasiveness from additional consensus quantity beyond a certain point (around 3 messages). This finding is consistent with the explanation captured by our Bayesian analysis, wherein increasing consensus quantity increases the amount of underlying evidence that the consensus represents but increases the probability of evidential overlaps between sources. This suggests that participants are doing more than simply counting up the number of sources in a consensus when assessing its persuasiveness. Interestingly, the point at which belief change plateaus in our data is at the same point noted by Shu and Carlson (2014), whose study suggests that there is a "charm of three", where persuasion is maximised at three messages, although we did not replicate the subsequent decline in persuasiveness that they reported at 4 messages.

Our results also indicate that expertise was a significant predictor of belief change, although with a small effect size. This finding aligns with Simmonds et al. (2023), where expertise was similarly identified as a significant predictor of belief change, but also with a small effect size. The replication of such a finding implies that while participants differentiate between sources based on their expertise, the extent to which they do is perhaps less than expected. A qualitative comparison between our experiment results and the Bayesian simulation indicate that participants perceived the expert organisation opinions as having only slightly higher evidential value than the non-expert opinions, which contrasts with previous findings that experts can be highly persuasive (e.g., Hornikx, Harris & Boekema, 2018; Jucks & Thon, 2017; Maddux & Rogers, 1980). Alternatively, our results may indicate that participants over-estimated the evidential value of the non-expert sources. Such an over-estimation could be explained based on various factors explored across the literature, including identity heuristics (Sundar, 2008), and the persuasiveness of anecdotal evidence (Zebregs, van den Putte, Neijens & de Graaf, 2015). Either explanation is concerning given our use of expert organisations, which each represent a consensus among individual experts. The similarity in persuasiveness between non-expert individual users and health organisations warrants further exploration.

Although our findings did not identify a significant interaction effect between expertise and consensus quantity, it should be noted that the relationship observed in Figure 4 does qualitatively resemble the relationship displayed in Figure 1, although with a much smaller effect of expertise. The most notable similarity is the plateauing of belief change that is most pronounced in the Expert condition. Akin to similar research (e.g., Connor Desai, Xie & Hayes, 2022; Simmonds et al., 2023), these findings further allude to differences in perceptions of source dependency between expert and non-expert sources. Further research would help elucidate possible effects.

It is important to note some possible limitations with our experiment which will inform future work. Firstly, the results from our attention check indicate minor discrepancies between the intended stance of our responses and the stance perceived by some of the participants. This discrepancy may be attributable to a lack of attention on the participants' behalf, but future research can strengthen the clarity of stances held by the responses.

Additionally, our sample of participants were primarily white males from the US, potentially making it difficult to generalise the findings to a broader population. As trust in institutions may vary across countries (Gil de Zúñiga, Ardèvol-Abreu, Diehl, Gómez Patiño & Liu, 2019), it is likely that the persuasiveness of expertise would vary. Gathering data from a more diverse sample would help improve generalisability.

Finally, we want to emphasize the preliminary nature of the Bayesian analysis we present here. It is based upon several simplifying assumptions, including the assumption that people have an idea about the finite amount of underlying evidence for/against a given claim, and that this is the primary factor influencing reasoning. We believe that this paper acts as strong initial point from which future research can investigate the reality of these assumptions.

Despite these limitations, our findings suggest several important implications. Consensus quantity is an important factor when aiming to change beliefs about health claims online – regardless of the expertise of those comprising the consensus. Concerningly, when the claim in question is false, the ability to sway opinion through pure consensus quantity can contribute to the spread of misinformation. However, we also note that by commenting on the veracity of online health claims, experts are also able to sway people in their favour. This implication reinforces the notion suggested in Simmonds et al. (2023) that experts can successfully act as influencers of public opinion in the face of misinformation. Encouraging health organisations to take an active role in online fact-checking and improving their visibility to the public will help people navigate the mass of health information available online. However, it is also important to note the diminishing returns that arise as consensus quantity grows (when there are no opposing messages) – a phenomenon that appears marginally more prevalent in response to experts. Thus, experts may be able to maximise their persuasive value if they choose to comment on health claims that remain unacknowledged by other experts.

Acknowledgements

This research was supported by a 2022 Digi+ FAME grant from The University of Adelaide.

References

Ahler, D. J., Roush, C. E., & Sood, G. (2021). The micro-task market for lemons: data quality on Amazon's Mechanical Turk. *Political Science Research and Methods*, 1–20.

Alister, M., Ransom, K. J., & Perfors, A. (2022). Source independence affects argument persuasiveness when the relevance is clear. *Proceedings of the Annual Meeting of the Cognitive Science Society, 44*(44).

Asch, S. E. (1951). Effects of group pressure upon the modification and distortion of judgments. In H. Guetzkow (Ed.), *Groups, leadership and men; research in human relations* (pp. 177–190). Carnegie Press.

Asch, S. E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological Monographs, 70*(9), 1–70.

Bovens, L., & Hartmann, S. (2004). *Bayesian epistemology.* OUP Oxford.

Chmielewski, M., & Kucker, S. C. (2020). An MTurk Crisis? Shifts in Data Quality and the Impact on Study Results. *Social Psychological and Personality Science, 11*(4), 464-473.

Connor Desai, S., Xie, B., & Hayes, B. K. (2022). Getting to the source of the illusion of consensus. *Cognition, 223*, 105023.

Fletcher, R., & Park, S. (2017). The Impact of Trust in the News Media on Online News Consumption and Participation. *Digital Journalism*, *5*(10), 1281–1299.

Gerard, H. B., Wilhelmy, R. A., & Conolley, E. S. (1968). Conformity and group size. *Journal of Personality and Social Psychology, 8*(1), 79–82.

Gil de Zúñiga, H., Ardèvol-Abreu, A., Diehl, T., Gómez Patiño, M., & Liu, J. H. (2019). Trust in Institutional Actors across 22 Countries. Examining Political, Science, and Media Trust Around the World. *Revista Latina de Comunicación Social*, *74*, 237-262.

Harkins, S. G., & Petty, R. E. (1981). Effects of source magnification of cognitive effort on attitudes: An information-processing view. *Journal of Personality and Social Psychology, 40*(3), 401–413.

Harkins, S. G., & Petty, R. E. (1987). Information utility and the multiple source effect. *Journal of Personality and Social Psychology, 52*(2), 260–268.

Hornikx, J., Harris, A. J. L., & Boekema, J. (2018). How many laypeople holding a popular opinion are needed to counter an expert opinion? *Thinking & Reasoning, 24*(1), 117–128.

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.

Kennedy, R., Clifford, S., Burleigh, T., Waggoner, P. D., Jewell, R., & Winter, N. J. G. (2020). The shape of and solutions to the MTurk quality crisis. *Political Science Research and Methods, 8*(4), 614–629.

Klein, G. (2008). Naturalistic Decision Making. *Human Factors, 50*(3), 456–460.

Leppink, J. (2019). *Statistical Methods for Experimental Research in Education and Psychology*. Springer International Publishing.

Lewandowsky, S., Ecker, U. K. H., & Cook, J. (2017). Beyond Misinformation: Understanding and Coping with the "Post-Truth" Era. *Journal of Applied Research in Memory and Cognition, 6*(4), 353–369.

Maddux, J. E., & Rogers, R. W. (1980). Effects of source expertness, physical attractiveness, and supporting arguments on persuasion: A case of brains over beauty. *Journal of Personality and Social Psychology, 39*(2), 235–244.

Mercier, H., & Morin, O. (2019). Majority rules: how good are we at aggregating convergent opinions? *Evolutionary Human Sciences, 1,* e6.

Nordholm, L. A. (1975). Effects of group size and stimulus ambiguity on conformity. *The Journal of Social Psychology, 97*(1), 123-130.

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* (pp. 833-839).

Shu, S. B., & Carlson, K. A. (2014). When Three Charms but Four Alarms: Identifying the Optimal Number of Claims in Persuasion Settings. *Journal of Marketing, 78*(1), 127–139.

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.

Stoica, P., & Selen, Y. (2004). Model-order selection: a review of information criterion rules. *IEEE Signal Processing Magazine, 21*(4), 36-47.

Suarez-Lledo, V., & Alvarez-Galvez, J. (2021). Prevalence of Health Misinformation on Social Media: Systematic Review. *Journal of Medical Internet Research, 23*(1), e17187–e17187.

Sundar, S. S. (2008). *The MAIN model: A heuristic approach to understanding technology effects on credibility* (pp. 73-100). Cambridge, MA: MacArthur Foundation Digital Media and Learning Initiative.

Whalen, A., Griffiths, T. L., & Buchsbaum, D. (2018). Sensitivity to Shared Information in Social Learning. *Cognitive Science, 42*(1), 168–187.

Zebregs, S., van den Putte, B., Neijens, P., & de Graaf, A. (2015). The differential impact of statistical and narrative evidence on beliefs, attitude, and intention: A meta-analysis. *Health communication, 30*(3), 282-289.

Zhao, S., Hu, S., Zhou, X., Song, S., Wang, Q., Zheng, H., Zhang, Y., & Hou, Z. (2023). The Prevalence, Features, Influencing Factors, and Solutions for COVID-19 Vaccine Misinformation: Systematic Review. *JMIR Public Health and Surveillance, 9*, e40201–e40201.