Examining People's Assumptions When Reasoning from Consensus

[redacted]

This thesis is submitted in partial fulfilment of the Honours Degree of Bachelor of Psychological Science

School of Psychology

University of Adelaide

September 2023

Word Count: 6998

Table of Contents

List of Figures

List of Tables

Abstract

Reasoning on social media is complex. Individuals face lots of information of varying accuracy from different sources, often encountering misinformation. To guide their judgment about claims, people may often turn to the opinions of others, relying on consensus cues. However, this reliance adds complexity because the quality of a consensus varies greatly, like the independence of evidence and sources. Despite this important variation, people's assumptions about the value of consensus quality information are unclear. To explore whether people prefer consensus information supported by different authors or the same author, and whether those author(s) provide different or the same reasons for their stance, individuals (N = 100) were recruited from Amazon's Mechanical Turk and exposed to claims on a mock social media platform. After rating their initial agreement with a claim, participants were shown diagrams summarising different combinations of diversity in authors and reasons in tweets responding to the claim. After selecting and reading their most preferred tweet, participants updated their agreement with the claim. Results showed systematic preferences across individuals. Consensus information involving different authors corroborating the same reason was preferred over all others. When this consensus type was unavailable, there were no systematic preferences overall. However, in these instances, there was systematicity within subgroups of individuals. Differences in consensus preferences did not lead to significant differences in agreement updating. These findings are important for understanding people's assumptions about consensus quality information and contribute to development of automated reasoning tools summarising consensus quality information for users to mitigate against misinformation.

Declaration

This thesis contains no material which has been accepted for the award of any other degree or diploma in any University, and, to the best of my knowledge, this thesis contains no material previously published except where due reference is made. I give permission for the digital version of this thesis to be made available on the web, via the University of Adelaide's digital thesis repository, the Library Search and through web search engines, unless permission has been granted by the School to restrict access for a period of time.

Signed

September 2023

Contributor Roles

ROLE	ROLE DESCRIPTION	STUDENT	SUPERVISOR 1	SUPERVISOR 2
CONCEPTUALIZATION	Ideas; formulation or evolution of overarching research goals and aims.	X	X	X
METHODOLOGY	Development or design of methodology; creation of models.	X	X	X
PROJECT ADMINISTRATION	Management and coordination responsibility for the research activity planning and execution.	X	X	
SUPERVISION	Oversight and leadership responsibility for the research activity planning and execution, including mentorship external to the core team.		X	X
RESOURCES	Provision of study materials, laboratory samples, instrumentation, computing resources, or other analysis tools.			
SOFTWARE	Programming, software development; designing computer programs; implementation of the computer code and supporting algorithms; testing of existing code.			X
INVESTIGATION	Conducting research - specifically performing experiments, or data/evidence collection.			X
VALIDATION	Verification of the overall replication/reproducibility of results/experiments.	X	X	X
DATA CURATION	Management activities to annotate (produce metadata), scrub data and maintain research data (including software code, where it is necessary for interpreting the data itself) for initial use and later re-use.			X
FORMAL ANALYSIS	Application of statistical, mathematical, computational, or other formal techniques to analyse or synthesize study data.	X		
VISUALIZATION	Visualization/data presentation of the results.	X		X
WRITING – ORIGINAL DRAFT	Specifically writing the initial draft.	X		
WRITING – REVIEW & EDITING	Critical review, commentary or revision of original draft		X	

Contribution Statement

This study was supported by a 2022 Collaborative Research Fund grant from Defence Innovation Partnerships, awarded to my supervisors.

Examining People's Assumptions when Reasoning from Consensus

Imagine you are scrolling through your social media feed when you come across a post claiming that "the U.S should not intervene in foreign disputes." How would you decide your level of agreement with this claim? You may not have direct access to relevant evidence and are unsure of the appropriate means of assessing the claim. In this kind of situation, you might use other means for evaluating your agreement, like noting other people's opinions on the claim and the prevailing consensus (the number of comments arguing that the U.S should not intervene). However, reliance on social evidence adds further complexity. The utility of a consensus is often contingent not only on the relative quantity of messages supporting versus rejecting a claim, but on consensus quality cues that might signal the independence of available evidence, like the number of different people sharing their opinion, and whether those people give the same or different reasons for their stance (Ransom et al., 2021). For example, a plausible-sounding reason corroborated by different people may be more compelling than a single person repeating themselves, and different people offering different reasons for their stance may be even more compelling. However, people often show limited sensitivity to these kinds of potentially important consensus quality cues (Ransom et al., 2021; Yousif et al., 2019). Therefore, the current study will directly examine people's assumptions about the value of such quality cues when reasoning from consensus. Specifically, it will examine whether people hold systematic preferences for a consensus derived from different authors or the same author, and whether those author(s) provide different or the same reasons for their stance. Ultimately, an understanding of people's assumptions about the value of these aspects of consensus quality will inform development of an automated reasoning tool that summarise useful consensus information for social media users.

Social Media and Reliance on Consensus

Many people receive information from social media: a third of U.S adults consume their news from Facebook (Pew Research Center, 2022). The structure of social media allows for fast sharing of information and lots of interactions. However, a downside is that users may experience 'information overload', being unable to evaluate all the content they face (Belabbes et al., 2022). This structure also makes social media a common tool for spreading misinformation (Allcott et al., 2019). Unfortunately, when users are subject to widespread misinformation, they may find it credible and adopt false beliefs (Lewandowsky et al. 2012).

Judging the truthfulness of claims in this environment is difficult. To help, users may look to the opinions of others. Indeed, research suggests that engaging in such social evaluation significantly impacts decision-making. In a classic conformity study, Asch (1956) had participants complete a perceptual task (judging the length of lines) after hearing answers from confederates who had been asked to provide incorrect answers. Many participants then gave an incorrect answer, conforming with consensus. The influence of consensus on attitudes is also evident online. Lewandowsky et al. (2019) asked participants to read blog posts endorsing or rejecting scientific consensus on climate change alongside comments replying to the post. The authors found a significant impact of consensus on attitudes: whenever the comments aligned with the position of the blog post, participants' agreement with the blog's argument increased (Lewandowsky et al., 2019).

Differences in Consensus Quality

Although a perceived majority has a significant impact on decision making and attitudes, appealing to consensus is not straightforward because there are potentially important qualitative aspects of consensus (Ransom et al., 2021). One way to highlight variation in consensus quality is to note the distinction between an independent and dependent consensus. To illustrate, imagine a new restaurant just opened in town. To decide whether you should go, you ask some friends who have already visited for their opinion. A relatively independent consensus would mean that each friend independently visited and gave their view of the restaurant to you. In contrast, if your friends only reference a single source (their mutual friend) who had told each of them that the restaurant is good, then the consensus is dependent. In this instance, the consensus is made up of repeated claims from one source (Mercier & Milton, 2019).

These differences in consensus quality can have significant implications: 80% of blogs denying global warming reference one primary source in their claims (Harvey et al., 2018). However, people are often insensitive to these differences. Yousif et al. (2019) conducted a study in which people read different news articles about a topic. Four articles took a positive stance, and one took a negative stance. Each news article cited a primary source, which was either identical across articles ("dependent consensus") or unique for each article ("independent consensus"). After reading the articles, participants indicated their confidence level in a related claim. Results showed that participants were not sensitive to the differences between an independent and dependent consensus, being similarly persuaded by both (Yousif et al., 2019). This kind of insensitivity has been observed by others including Ransom et al., (2021) who found belief revision was impacted primarily by the quantity of posts supporting versus opposing the claim, rather than markers of consensus quality (e.g., evidence diversity). Similarly, Simmonds et al., (2023) found belief in health claims was primarily impacted by consensus quantity with minimal consensus quality effects (e.g., source diversity).

Exploring the source of this insensitivity, Connor Desai et al. (2022) tested whether people's uncertainty about the independence of primary sources is a contributing factor. After replicating the key finding of Yousif et al. (2019) the authors took steps to highlight the independence of sources, resulting in participants giving more weight to a claim based on an independent consensus (Connor Desai et al., 2022). Furthermore, people also exhibit increased sensitivity to the independence of primary sources when it is made clear to them that individuals are actually reasoning based on the primary sources (Alister et al., 2022). These findings suggest that people *can* incorporate cues to consensus quality when assisted.

Current Social Media Reasoning Tools

The insensitivity to consensus quality identified by Yousif et al. (2019) and Ransom et al., (2021) suggests people may require assistance in considering this information online. Tools providing such information could complement other initiatives in mitigating against misinformation. Current tools like warning labels notify users of information that may be misleading via a third-party fact checker. Whilst this may be effective for identifying misinformation, the process of determining the legitimacy of content is slower than the rate at which misinformation spreads, meaning only a small proportion of misinformation will have labels attached (Pennycook et al., 2020). Another approach is crowd-sourced judgments, drawing from the 'wisdom of the crowds' literature showing that aggregated individual judgments form accurate overall judgments (Epstein et al., 2020). Crowdsourced trust ratings can accurately identify mainstream news sources and fake news sources (Pennycook & Rand, 2019), but implementing crowdsourced judgments on social media may have partisanship issues (Allen et al., 2022). These strategies are also limited in cases that cannot simply be fact-checked (e.g., for considering whether "the U.S should intervene in foreign disputes").

As another possible tool to support people in online environments, researchers have begun developing an automated reasoning aid summarising consensus quantity and quality cues (Le Leu, 2021; Ransom & Stephens, 2023). One study found that a reasoning aid visually summarising consensus information helped people to consider quality cues of authors and reason diversity (Le Leu, 2021). Therefore, such a tool could be helpful in conjunction with current counter-misinformation strategies to aid social media users.

The Ground Truth Problem for Consensus-Based Reasoning Tools

In developing a consensus-based reasoning tool and assessing its effectiveness in informing people's beliefs, an important challenge is that there is no single ground truth for deciding the value of a particular kind of consensus. The research by Yousif et al. (2019), for instance, presupposes that information from an independent consensus is more valuable. Although this is often true, a dependent consensus could be advantageous in particular contexts. For example, if a source has expertise or a more reliable method for investigating a claim, perhaps one should attach more value to repeated reports from this dependent source than independent sources without these characteristics (Connor Desai et al., 2022). There are also instances in which connected (dependent) sources produce more reliable information than independent sources. Pilditch et al. (2020) demonstrated that when opinions are contradictory (serving to disconfirm a previous opinion) dependent evidence can be advantageous compared to independent evidence (Pilditch et al., 2020; Connor Desai et al., 2022). Given that either independent or dependent evidence can be more informative in various contexts, unfortunately one cannot simply assume that a reasoning tool is effective if it routinely encourages people to be more influenced by an independent consensus. An alternative approach for the tool could be to ask reasoners which kinds of consensus information they value, then tailor the tool to their preferences. However, we first need to investigate whether people have particular preferences for consensus information.

Consensus Quality Preferences

Although research has demonstrated people's limited sensitivity to consensus quality as measured by the impact on belief (e.g., Ransom et al., 2021) there has been little direct investigation into people's assumptions about the value of independent or dependent evidence. For instance, do people find it more valuable to see a single author giving different reasons for their stance, or independent authors corroborating a reason? One study on sensitivity to evidential dependencies found variation in people's a-priori beliefs about the value of different evidence (Xie & Hayes, 2022). Xie and Hayes found that 38.5% of participants reported that independent information (i.e., from separate sources with independent observations) is more useful, 36.5% selected dependent information (sources who had also seen evidence from other people), and 25% had no preference. When asked for justification, the most common theme amongst independent information endorsers was the biases created by informational dependency. In contrast, dependent endorsers commonly referenced the accumulative, sequential nature of evidence. Similarly, results from Experiment 4 of Yousif et al. (2019) found that when asked directly, many participants believed an independent consensus to be important, with 50% of participants believing that, in a hypothetical scenario, news articles citing independent sources would be more believable (Yousif et al., 2019).

The broader literature suggests people's assessment of the value of consensus information may be based on heuristic cues like message characteristics (e.g., repetition or different arguments) or source characteristics like repeated exposure (Chaiken, 1987).

Supporting that repetition from one source may be valued, Weisbuch et al., (2003) found that prior exposure to a source increased participants' agreement with messages from that source. Alternatively, reasoners may value the presence of different sources in a consensus: different sources providing different arguments in favour of a claim has been found to be more persuasive than one source providing arguments (Harkins & Petty, 1981). Similarly, the characteristics of messages themselves may be valued differently. Repeated presentation of an argument has been found to be more persuasive than a single presentation (Weaver et al., 2007), and repeated presentation of social media posts increases their perceived truth (Nadarevic et al., 2020). Alternatively, people may infer that different arguments are valuable. For example, the persuasiveness of testimony has been found to be a positive function of the number of different arguments a source offers (Calder et al., 1974). This conflicting evidence indicates that people could value a repeated source or argument in a consensus but could also find a consensus with different sources or arguments compelling. Reasoners may also make inferences about the value of combinations of both source and message characteristics. For example, a source able to provide multiple different arguments rather than repeating one argument may be thought to possess more expertise and therefore be useful, since people assume experts are more knowledgeable than non-experts (Vaupotic et al., 2021).

Current Study

The aim of the current study is to examine people's assumptions when reasoning from consensus. A primary motivation is also to provide insights for future development of reasoning tools providing consensus quality information to users. I will investigate whether people prefer to read a message about a claim with: 1) a reason that is shared by either different authors or the same author repeatedly and 2) a reason that is provided multiple times or only once. In a novel experiment on a mock social media platform, after indicating their initial level of agreement with a claim, participants will be shown a diagram summarising consensus quality (author and reason diversity) in tweets supporting the claim (with the diagram based on Le Leu, 2021; Ransom & Stephens, 2023). Each diagram will describe two different combinations or 'motifs' of author and reason diversity. These motifs will be constructed using 2x2 factorial combinations, resulting in: different authors giving different reasons, different authors each repeating the same reason, the same author giving different reasons, and the same author repeating the same reason. Hence, there will be six different diagrams, each with two of these motifs per diagram. Participants will be asked to choose which tweet they would most like to read based on the diagram. After reading the tweet, they will indicate their updated agreement level with the claim.

The study will test several key hypotheses related to participants' choices from the six different pairwise comparisons of authors and reasons in the six different diagrams. Note that the order of the following hypotheses corresponds to systematic diagram numbering (diagram 1 to 6; see Method), rather than numeric order of hypotheses themselves.

First, a source offering different arguments for their position is persuasive, and people may infer that a source providing different reasons for their stance has expertise, with perceived expertise influencing the perceived truth of statements (Calder et al., 1974; Nadarevic et al., 2020). Hence, people may infer that a source providing different reasons for their stance is worth hearing from compared to different authors providing one reason each. For Diagram 1, it is hypothesised that:

H1 Participants will systematically prefer to read tweets from the same author giving different reasons over different authors each giving a different reason.

Additionally, when asked directly, people often indicate that evidence derived from independent sources is more valuable than dependent sources (Xie & Hayes., 2022; Yousif et al., 2019). The prevalence of a position also impacts attitudes (Lewandowsky et al., 2019) Therefore, people should prefer evidence corroborated by different sources. For Diagrams 2, 4 and 6 it is hypothesised that:

H2 Participants will systematically prefer to read tweets from different authors giving the same reason over the same author giving different reasons.

H4 Participants will systematically prefer to read tweets from different authors giving the same reason over different authors each giving a different reason.

H6 Participants will systematically prefer to read tweets from different authors giving the same reason over the same author repeating the same reason.

The repetition of statements significantly impacts people's perception of truth (Nadarevic et al., 2020; Weaver et al., 2007). Therefore, in the context of reasoning from consensus, people may infer that a source repeating the same reason is more compelling than different sources each giving a reason once. In Diagram 5 it is hypothesised that:

H5 Participants will systematically prefer to read tweets from the same author repeating the same reason over different authors each giving a different reason.

Finally, as noted above, the persuasiveness of testimony from a source has been shown to be a function of the number of different arguments the source offers (Calder et al., 1974). When multiple arguments are offered, people may infer that the source has more expertise in the topic, with source credibility impacting the perceived truth of statements (Nadarevic et al., 2020). However, repeated arguments have also been found to be a cue affecting attitudes (e.g., Weaver et al., 2007). Given this conflicting evidence, it will be interesting to see if people have a preference, but there will be no specific prediction for Diagram 3 comparing the same author giving different reasons against the same author repeating the same reason.

Although not of primary interest, the study will also examine whether there are significant differences in agreement change (initial to updated) between those who chose opposing motifs in each diagram. This analysis will assess whether different preferences in consensus information affect change in agreement with the claims in case some kinds of preferred tweet types are more persuasive.

Method

To examine people's assumptions when reasoning from consensus, I conducted the following online experiment. On each trial, participants were presented with a diagram summarising both the diversity of authors tweeting in response to a claim and the diversity of reasons presented in the tweets. I collected participants' choice of tweet to determine their preference for reading tweets shared by either different authors or the same author with a reason that is provided multiple times or only once. I also collected participants' agreement rating with the claims.

Pre-Registration

I completed a pre-registration before data collection. This included the hypotheses, variables, target sample size, and planned main analyses (see Appendix A).

Participants

Before providing consent, participants were informed of the general aims and methods of the study and informed that they could withdraw at any point. Respondents' data was collected using anonymous ID numbers. The research was approved by the Human Research Ethics Committee, School of Psychology, University of Adelaide (Ref. 22/75).

There were 100 participants recruited through Amazon Mechanical Turk. Participation took 20-30 minutes (including instructions), and participants were compensated 3.00 USD. Inclusion criteria required participants to be 18 years or older and be proficient in English (which was pre-screened). The age of the sample ranged from 21 to 76 years (mean age 38.42 years) comprising 60% males, 37% females, and 3% other. The sample was mostly from the U.S and Brazil (92%) and most participants were White (64%), with the remainder identifying as Asian (10%), Black (8%), Latinx (8%), multi-racial (5%), and Indian (3%). Most participants used social media daily (86%), were politically left-wing (57%), and were native English speakers (80%). All had a minimum equivalent of a high-school qualification.

Design

Using a discrete choice experiment design, the number of authors and reasons were both 2-level factors that varied systematically within the motifs in the diagrams (same vs. different authors and same vs. different reasons). The levels of the author diversity factor were factorially paired with the levels of the reason diversity factor to create four motifs summarising the different possible combinations of diversity of authors tweeting and the diversity of reasons presented in the tweets (see Figure 1). In each trial, the diagram showed a pairwise contest between two different consensus motif choices. There were six different diagrams based on the six pairwise motif combinations (see Figure 2). These diagrams varied within- subjects and between- trials. Participants saw each diagram three times, forming 18 trials.

Materials

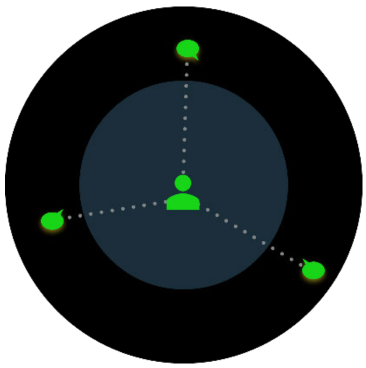
Diagrams

In the diagrams, within each motif, tweets responding to a claim came from either the same author*,* who tweeted three times, or different authors*,* who each tweeted once. Authors were designated by person icons within the inner circle of the diagram. Diversity of reasons was shown in the outer circle, illustrated by speech bubbles. The same reasons were those in which the same underlying reason was provided in three different tweets, indicated by speech bubbles clustered together. Different reasons were those in which a distinct reason was provided in a tweet, illustrated by speech bubbles far apart. Figure 1 and 2 illustrates these motifs and pairwise combinations.

Figure 1.

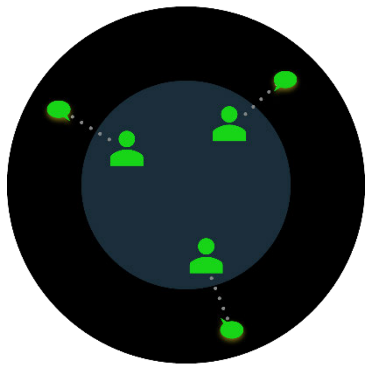
The Four Motifs Used in the Study.

Motif 1



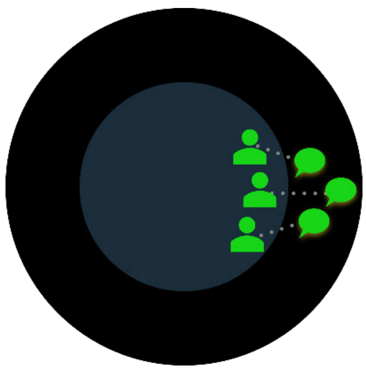
The same author giving different reasons.

Motif 2



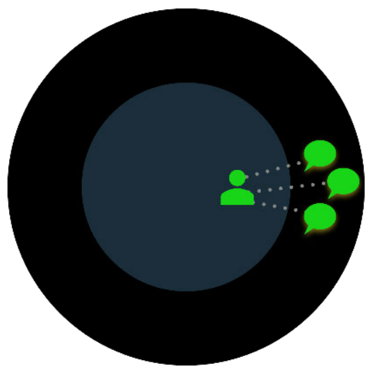
Different authors each giving a different reason.

Motif 3



Different authors each giving the same reason.

Motif 4

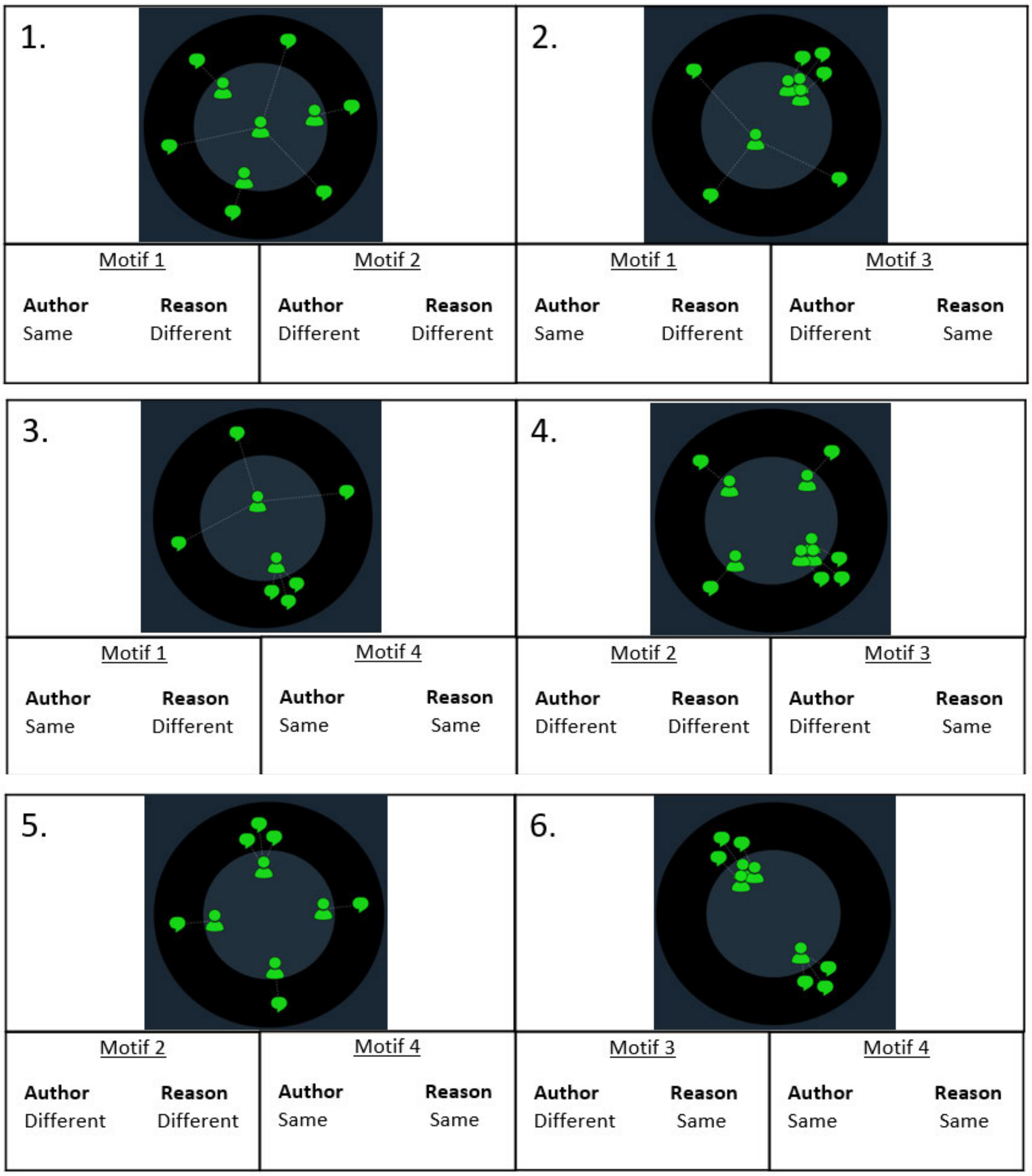


The same author giving the same reason.

Note. Person icons indicate number of authors tweeting. Speech bubbles indicate the number of reasons in the tweets. Bubbles clustered together indicate the same reason and bubbles spaced apart indicate different reasons.

Figure 2.

Design of the Six Diagrams Presented During the Experiment.



Note. The six diagrams (labelled 1 to 6) each contain two motifs. The corresponding authors and reasons in the motifs are described below each diagram.

Claims

Before the main experiment, I conducted a pilot study using 36 claims, the results of which informed selection of claims with a range of initial agreement distributions for the main experiment to avoid extreme polarisation. Using Amazon Mechanical Turk, I presented participants (*N* = 59) with a broad range of claims. After reading the claim, participants were asked to rate their agreement on a scale from -50 (“strongly disagree”) to 50 (“strongly agree”). A range of claims with positive and negative skew and normal distributions were incorporated into the main experiment.

The 18 claims presented to participants with the diagrams are displayed in Table 1. Five claims were adopted from the topics in Ransom et al. (2021) with minor edits. The thirteen remaining claims were taken from the pilot data. The claims were intended to elicit a range of initial agreement ratings across various topics. For example, claims were defence related, had a scientific nature, appealed to personal experiences, or were opinion-based. Claims were randomly allocated to diagrams across the 18 trials. Each claim was shown in a neutral format without context (see Figure 3). The language of the claims was simple to make them accessible to participants.

Tweets

Each trial included a tweet arguing in favour of the claim. Although there were six tweets that participants could choose between, they were shown the same tweet regardless of tweet choice because they were choosing only one tweet per trial. Tweet stance was constant across all trials (pro). Table 2 shows examples of a tweet across different claims. Tweets used plain language and usernames of the authors and user images were randomly generated.

Table 1.

The Eighteen Claims Used.

#	Claim
1.	Perfect avocados are getting harder to find.
2.	Sodium-ion batteries will replace lithium-ion.
3.	China will increase its trade sanctions against Australia.
4.	Charitable giving will increase over the next three years.
5.	The government should reduce spending of foreign aid.
6.	Australia should not acquire nuclear submarines.
7.	The U.S should not intervene in foreign disputes.
8.	A college degree is not worth it.
9.	Ukraine should cede territory to end its war with Russia.
10.	Paper money and coins should be phased out.
11.	Genetically modified crops are a bad idea.
12.	National service should be mandatory.
13.	Working from home is more productive than in the workplace.
14.	Ukraine should be allowed to join NATO.
15.	Social media is unsafe for children.
16.	Tensions in the Indo-Pacific are on the rise.
17.	Children learn better by handwriting than by typing.
18.	Crypto currencies should be tightly regulated.

Figure 3.

Screenshot of a Claim and Agreement Scale.

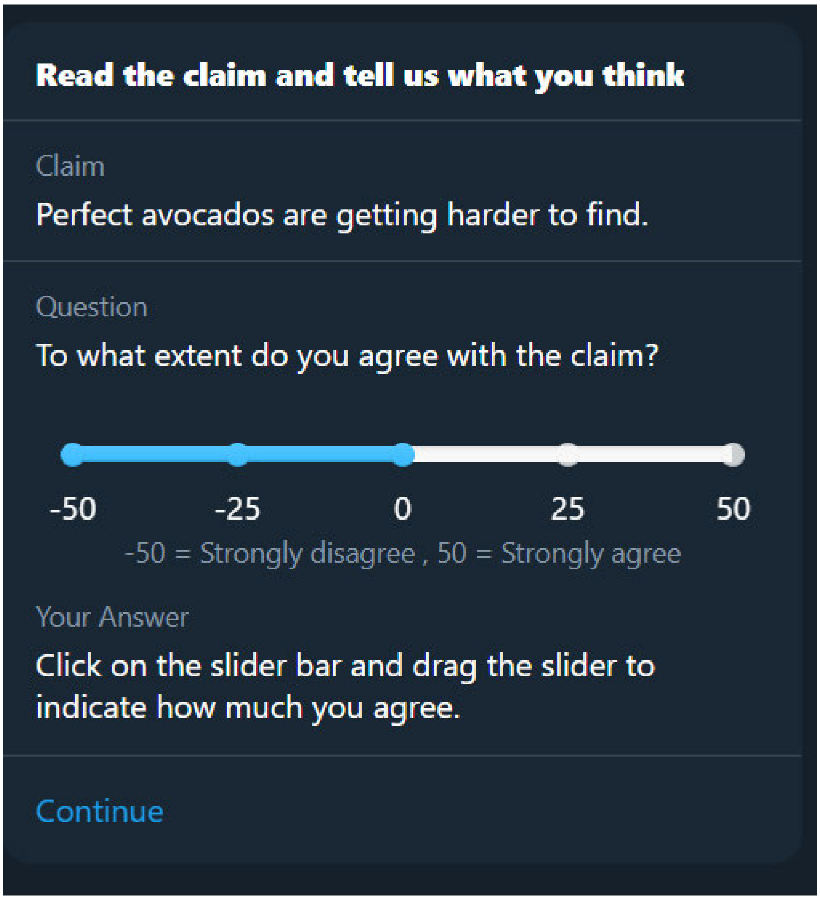


Table 2.

Example Tweets in Response to Claims.

Claim	Tweet
Perfect avocados are getting harder to find.	As soil becomes depleted through over farming, there are less nutrients available for plants like avocados which makes it harder for farmer to get their trees bearing great fruits.
Crypto currencies should be tightly regulated.	Strict crypto regulations are crucial in preventing money laundering, crime, and illicit activities.
Working from home is more productive than in the workplace.	Working from home reduces risk of illness because it limits contact with others. More productive and healthier!
Ukraine should be allowed to join NATO.	Ukraine should be allowed to join NATO: the country has actively pursued regional security against Russian aggression, strengthening confidence for its allies in the region.

Dependent Variables

I used two dependent measures. The first was participants choice of tweet to read after viewing the diagram. This was used to record preferences for the number of authors and reasons in a consensus. Selecting one of the available tweets required participants to make a discrete choice between one consensus motif or the other. The second measure was participants' agreement rating with a claim on a scale from -50 ("strongly disagree") to 50 ("strongly agree"). This was recorded before and after: viewing the diagram, choosing a tweet, and reading the selected tweet. The measure was used to assess any impact of tweet selection from the diagram on agreement change. Agreement ratings were also included to encourage participants to think carefully about a claim and their tweet choices.

Procedure

After giving informed consent, participants were asked to provide their demographic information. Next, participants were provided with task instructions and detailed explanations of the motifs and diagrams. To ensure they understood the instructions and diagrams, participants were required to correctly answer multiple-choice questions before beginning the experiment (see Appendix B).

In each trial, participants were presented with a claim on a mock social media platform. Participants were asked to read the claim and indicate their level of agreement by moving the slider bar on a scale from -50 ("strongly disagree") to 50 ("strongly agree"). Figure 3 displays this step. Next, participants were presented with a diagram summarising the number of authors and reasons in tweets responding to the claim. The diagram was presented alongside panels signalling the six available tweets, with only the names and images of authors shown (see Figure 4). If participants hovered their cursor over a speech bubble in the diagram, the corresponding tweet panel was highlighted (and vice versa). The arrangement of authors and reasons in the tweets were represented by two discrete motif choices in the diagram. Participants were asked to report which tweet they would most like to read by selecting one of the speech bubbles. Clicking on a speech bubble revealed the tweet text, which participants were asked to read (see Figure 5). Participants were then given chance to update their agreement on the slider bar (see Figure 5). This procedure was repeated for all trials, with trials shown in random order for each participant. Upon experiment completion, participants were debriefed and informed that the opinions of the researchers, research institutions and funding bodies involved were not reflected in the claims or tweets, and the intention was not to promote any point of view. Rather, researchers were interested in people's beliefs about different topics often discussed on social media.

Figure 4.

Screenshot of Experiment During Tweet Selection.

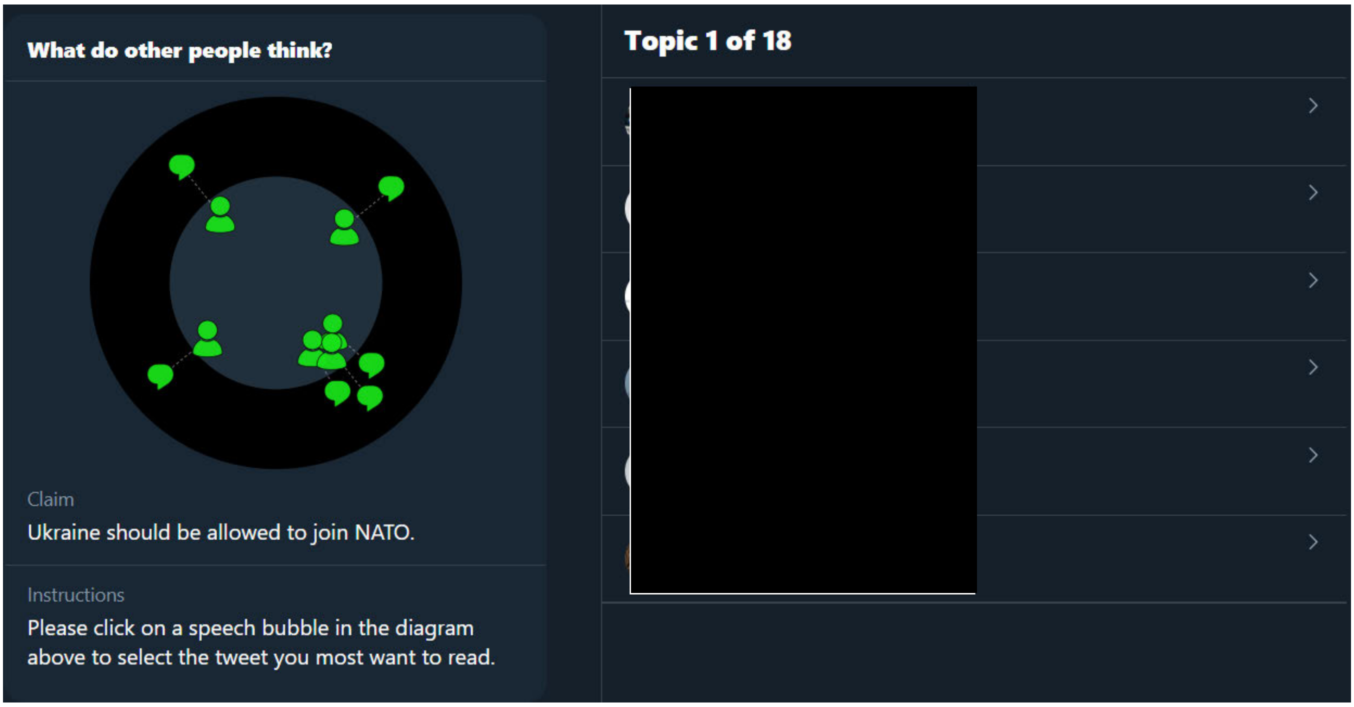


Figure 5.

Screenshot of Selected Tweet.

Results

Initial Agreement Ratings

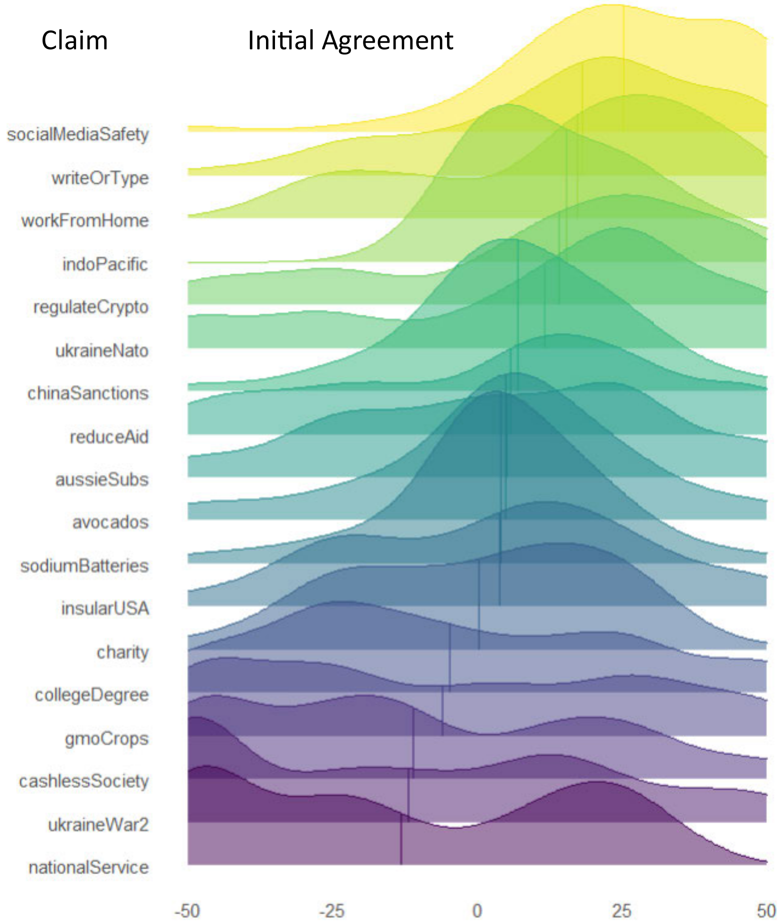
Figure 6 displays distributions of initial agreement ratings with the claims prior to viewing the diagram. Skewness values ranged between -1 to 1, indicating that distributions were not extreme. As Figure 6 illustrates, there was a range of initial agreement ratings across claims and participants. Therefore, the study's hypotheses were tested within a context of a variety of initial agreement ratings, across various topics.

Change in Agreement

To test for any impact of motif choice on agreement, I compared the mean difference (delta) in agreement ratings (initial to updated) between participants who chose different motifs in each diagram. This analysis was exploratory, and no specific predictions had been made. Since each participant was receiving the tweet that they most wanted to read before updating agreement, there were not necessarily any expected significant differences, but some motifs might have been more persuasive than their opposing option. In each case, delta was not significantly different between the two groups (those who chose opposing motifs in each diagram; $p > .05$ based on independent samples t-tests). Appendix C reports results of the t-tests for each diagram (Table C1) and box plots of differences in mean agreement change between motif choices (Figure C1). The overall mean change in agreement (delta) across the six diagrams was 6.06 with the overall standard deviation of 11.94.

Figure 6.

Distribution of Initial Agreement Ratings.



Note. Density plots indicating the range of initial agreements with claims from -50 ("strongly disagree") to 50 ("strongly agree").

Consensus Quality Preferences

Addressing the key hypotheses, I conducted a series of binomial tests to examine whether there was a significant preference for one motif over the other across participants, for the six diagrams. Table 3 displays the results of the binomial tests and Figure 7 displays the six diagrams (with the different motifs colour-coded; column 1), the corresponding motif preferences across participants (column 2), and preferences within individuals (column 3). As reflected in column 2, the binomial tests indicated that motif counts significantly deviated from chance in Diagrams 2, 4, and 6. As hypothesised in *H2, H4,* and *H6*, across participants the proportion of counts for preferring different authors giving the same reason was significantly higher than for the same author giving different reasons (Diagram 2), different authors each giving a different reason (Diagram 4), and the same author giving the same reason (Diagram 6). In all other cases, where different authors giving the same reason was not in a diagram, binomial tests indicated that the proportion of counts across participants did not significantly deviate from chance. This is reflected in Figure 7 (column 2) for Diagram 1, 3 and 5. This meant that hypothesis *H1*, predicting that participants would systematically prefer tweets from the same author giving different reasons, was not supported. Hypothesis *H5*, predicting that participants would systematically prefer to read tweets from the same author repeating the same reason, was also not supported. No prediction was made for Diagram 3, and the non-significant binomial test result is perhaps unsurprising: tweets from both the same author giving different reasons and the same author repeating the same reason could be seen as compelling.

Figure 7.

Consensus Quality Preferences.

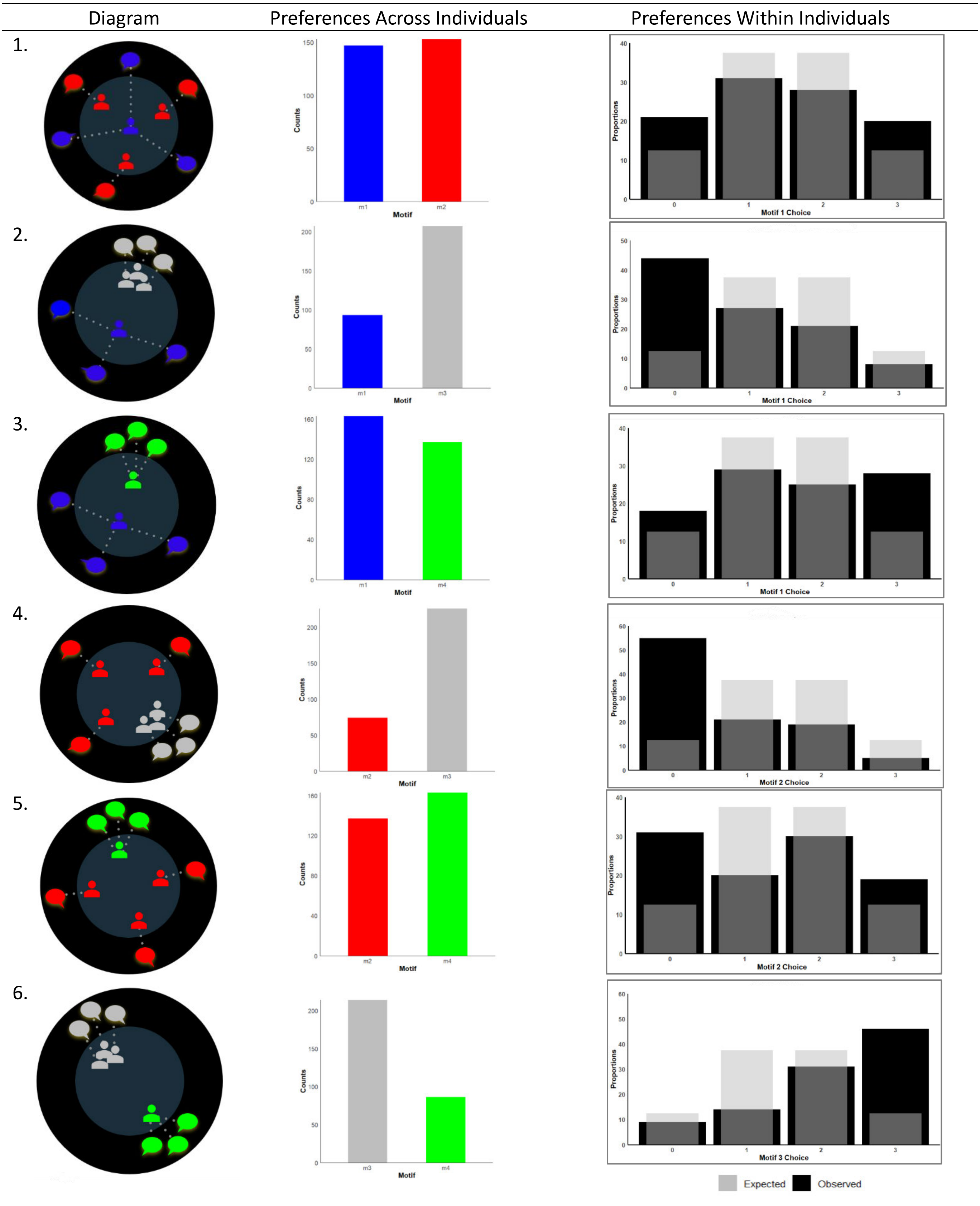


Table 3.

Binomial Tests of Equal Proportions for Overall Motif Choices.

Diagram	N. of successes	Total N. of observations	Significance
1	147	300	0.77
2	93	300	<.001
3	163	300	0.14
4	74	300	<.001
5	137	300	0.14
6	214	300	<.001

Note. Probability of success in each diagram = 0.5.

As an exploratory analysis, I also examined whether group-level preferences were reflected at the level of individual participants, across trials. This analysis is important because, for example, no preference across participants could either reflect that individuals generally have no preference, or that different subgroup(s) have opposing preferences. Column 3 in Figure 7 shows both the observed (black) and expected (grey) frequencies of choosing a motif zero, once, twice, or three times from the three times each diagram was presented across the 18 trials. The expected frequencies were determined with a binomial distribution simulation, capturing random motif choices (probability of 0.5).

To test whether individuals chose a motif zero, once, twice, or three times at levels different from chance, a chi-square goodness-of-fit test was used for each diagram (see Table 4). The observed frequency of choice was significantly different from the expected frequency across all six diagrams. The overall preference for different authors giving the same reason seemed to be reflected at the individual level for most participants (Diagrams

2, 4, and 6 in column 3 of Figure 7). In Diagram 1, where there was no group-level preference, there was possibly subgroups of individuals who held systematic preferences for the same author giving different reasons (motif 1) or different authors each giving a different reason (motif 2). Specifically, the proportions of either motif being chosen zero times or three times out of three that Diagram 1 was presented seemed above chance levels (see Figure 7). Similarly, in Diagram 3 and 5, where there were no clear group-level preferences, there seemed to be evidence of systematicity within individuals. In Diagram 3, the proportions of either the same author giving different reasons (motif 1) or the same author repeating the same reason (motif 4) being chosen zero times or three times appeared to be above chance levels. In Diagram 5, the proportions of either different authors giving different reasons (motif 2) or the same author repeating the same reason (motif 4) being chosen zero times or three times also exceeded chance levels. These results indicate that instances in which there was no clear group-level preference for a motif did not mean participants generally had no preferences. Rather, different subgroups may have had systematic preferences for different motifs in these diagrams.

Table 4.

Chi-Square Goodness-of-fit Tests for Individual-Level Consistency in Motif Choice.

Diagram	Chi-Square	Degrees of Freedom	Significance
1	13.81	3	.003
2	91.20	3	<.001
3	27.73	3	<.001
4	165.39	3	<.001
5	40.42	3	<.001
6	106.61	3	<.001

Discussion

Summary of Findings

The current study aimed to examine people's assumptions when reasoning from consensus. I explored whether people hold systematic preferences for a consensus derived from different authors or the same authors, and whether those authors provide different or the same reasons for their stance. To my knowledge, the current study is the first to use visual summaries of consensus quality information to directly explore these assumptions. The study used a mock social media interface, extending on the work of Ransom et al. (2021), which allowed for investigation of consensus quality preferences in a realistic social media context with various claims. The key finding was that people systematically preferred a consensus derived from different authors corroborating the same reason. This motif was preferred by most people to all other motifs, supporting hypotheses *H2, H4,* and *H6.* When this motif was not available, there was no overall preference. This meant that hypotheses *H1,* predicting that people would systematically prefer a consensus made up of the same author giving different reasons, was not supported. Similarly, hypothesis *H5*, predicting that people would systematically prefer a consensus derived from the same author giving the same reason, was not supported. There was also no overall preference for either the same author repeating the same reason or the same author giving different reasons in Diagram 3. Further individual-level analysis revealed that in diagrams where there was no overall preference, there appeared to be systematicity within subgroups of individuals. Change in agreement with the associated claims was not significantly different for those who chose opposing motifs across the six diagrams.

Comparison of Findings with Previous Literature

A consensus derived from different authors giving the same reason was systematically preferred over the same author giving different reasons (Diagram 2) and the same author repeating the same reason (Diagram 6), supporting hypotheses *H2* and *H6*. This suggests that, regardless of the number of reasons a single author offers (repeated vs. different) for their stance, people still find the presence of different authors more informative. These findings align with literature showing that when asked directly, people often identify evidence derived from independent sources as more valuable a-priori (e.g., Xie & Hayes, 2022; Yousif et al., 2019). Importantly, however, this overall preference in the current study was contingent on those different authors all giving the same reason for their stance. In comparison, when three different authors each gave a different reason there was no overall preference for this motif (e.g., Diagram 1). The clearest instance of this was found in the results of Diagram 4: there was an overall preference for different authors giving the same reason over different authors each giving a different reason, supporting hypothesis *H4.* This differs from previous findings in the consensus literature indicating that people are often insensitive to the consensus quality cue of reason diversity. Ransom et al. (2021), for example, found no effect of reason diversity on revised belief in claims. The key finding in the current study of a group-level preference for different authors giving the same reason in the current study also runs in contrast to broader literature indicating that, when individuals are aware a message source has a persuasive motive, a source making their case with three different endorsements has a greater positive impression than one or two endorsements (Shu & Carlson, 2014). Overall, this finding suggests that people infer that the same reason being repeated by different authors is a clue to its usefulness.

In comparison, when different authors giving the same reason was not available in a diagram, there were no systematic preferences overall. The same author giving either different reasons (Diagram 1) or repeating the same reason (Diagram 5) was not preferred across participants to a consensus of different authors each giving a different reason, failing to support hypothesis *H1* and *H5.* However, at the individual level, different subgroups seemed to have opposing preferences. These subgroups suggest that individuals may differ in their inferences about the value of the same author giving either different (Diagram 1) or the same reasons (Diagram 5) compared to different authors each giving a different reason for their stance. In Diagram 1, for example, it was hypothesised that people would assign more weight to the same author giving different reasons. This was predicted because a source offering different arguments for their position is persuasive, and people may infer that a source capable of providing different reasons for their stance has expertise, which influences the perceived truth of statements (Calder et al., 1974; Nadarevic et al., 2020). This individual variation also highlights the problem of no ground truth for deciding the value of a consensus: individuals may have made opposing, yet equally valid inferences about the contexts in which a consensus derived from the same author or different authors would be advantageous.

As an interesting finding from Diagram 3, there was no overall preference for either the same author giving different reasons or the same author repeating the same reason. No prediction had been made for this diagram based on prior conflicting evidence that both repeated and different arguments from a source impact persuasiveness (Calder et al., 1974; Weaver et al., 2007). In the current study, the individual-level analysis revealed that some individuals may have had systematic preferences for the same author giving either different reasons or repeating the same reason. These individual differences may be a factor in the conflicting literature on the persuasiveness of repeated and different arguments from a source. There is little work in the consensus literature investigating preferences for a dependent source providing either different reasons or the same reason, but the Diagram 3 results in the current study may be explained elsewhere. Evidence in the explanatory preferences literature, for example, suggests that individuals may differ in how satisfying they find simple or complex explanations. On the one hand, people intuit that a simple, broad explanation is indicative of a high-quality explanation, viewing simplicity as an explanatory virtue (Lombrozo, 2016). On the other hand, Zemla et al. (2017) found that people prefer complex explanations invoking several causal mechanisms. The simple explanation is more like a single author repeating one compelling reason, which may be preferred by individuals for its simplicity and breadth, whilst the complex explanation is like the same author giving different reasons, equivalent to an explanation invoking different causal mechanisms to explain an effect (Zemla et al., 2017).

Applied Implications

The findings of the current study have significant implications for understanding and supporting reasoning from consensus on social media. The rise of misinformation on social media has become a global concern (Urakami et al., 2022), and failure to attend to consensus quality online can have significant implications (e.g., Harvey et al; 2018). Unfortunately, there is evidence that people are often insensitive to consensus quality cues when reasoning (e.g., Ransom et al., 2021). However, recent work suggests that people *can* incorporate consensus quality information into their reasoning strategies with assistance (e.g., Connor Desai et al., 2022), and researchers have already begun developing automated reasoning tools summarising consensus quality cues (e.g., Ransom & Stephens, 2023). The current study will contribute to the design of such tools, establishing the assumptions people have about the value of consensus quality information in a realistic online environment. Given the value of consensus quality information can be context dependent (e.g., Pilditch et al., 2020), a reasoning tool tailored to group or even individual preferences could shine a light on aspects of consensus quality information that are relevant in a scenario but would otherwise not be attended to by reasoners because they do not align with their assumptions. Using a tool to summarise this information could encourage users to consider these aspects when reasoning about claims on social media and will be helpful alongside current counter-misinformation strategies.

Limitations and Future Directions

There were some potential limitations in the current study that future research could address. First, the diagram used had previously been shown to be effective (Le Leu, 2021; Ransom & Stephens, 2023) but may not necessarily have been the most optimal for conveying consensus quality information. The diagrams used principles of multi-dimensional scaling, like distance to convey reason diversity (Ajjour et al., 2018). Although this allowed for summaries of diversity of authors and reasons in the current study, there may be scalability issues if researchers wanted to convey lots of combinations of authors and reasons in a diagram. People's reasoning strategies in situations of uncertainty also depend on the visualisation they are presented with (Eichler et al., 2020). Therefore, people's perceptions of the value of consensus motifs may have been dependent on the visualisation used. Future research could test different presentations of consensus quality information and observe whether this impacts preferences.

Although not a primary outcome measure, another limitation was the method of measuring agreement change. The requirement for participants to update agreement immediately after choosing a tweet could have led to demand effects, encouraging shifts in agreement. Therefore, the degree of agreement change may be overestimated in the current study. More importantly, the current study required participants to choose only one tweet per trial. This method was intended to elicit preferences by having people prioritise one tweet. However, choosing only one tweet may have encouraged arbitrary responses rather than reasoning about the value of consensus quality cues. A future study could require participants to select all available tweets in a diagram from most to least preferred to test for preference consistency beyond first tweet choice. Finally, to further explicate the individual differences in preferences, a future study could ask participants to give written justifications for their preferred consensus motif. Key themes could be collated and examined to see how the justifications relate to the number of authors and reasons in the chosen motif.

Conclusion

The current study examined people's assumptions when reasoning from consensus. It sought to directly explore people's preferences for consensus quality information like the number of authors and reasons in a consensus in a social media context. Results showed that people do hold some assumptions about consensus quality information: most systematically preferred a consensus derived from different authors corroborating the same reason for a position. This pairing of authors and reasons was favoured over all other consensuses. When this option was not available, there were no overall preferences. However, even in these instances, there was still evidence of systematicity for opposing motif preferences within different subgroups of individuals. Ultimately, these findings may inform the development of a robust, automated reasoning tool summarising helpful consensus quality information for social media users to combat misinformation online.

References

Ajjour, Y., Wachsmuth, H., Kiesel, D., Riehmann, P., Fan, F., Castiglia, G., Adejoh, R., FrÖhlich, B., & Stein, B. (2018). Visualization of the topic space of argument search results in args.me. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 60-65

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), 2767-2773.

Allcott, H., Gentzkow, M., & Yu, C. (2019). Trends in the diffusion of misinformation on social media. *Research & Politics, 6*(2), 1-8. https://doi.org/10.1177/2053168019848554

Allen, J., Martel., C., & Rand, D. G. (2022). Birds of a feather don't fact-check each other: Partisanship and the evaluation of news in Twitter's Birdwatch crowdsourced fact-checking program. *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, *245,* 1-19. https://doi.org/10.1145/3491102.3502040

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.

Belabbes, M. A., Ruthven, I., Moshfeghi, Y., & Rasmussen, P. D. (2023). Information Overload: A Concept Analysis. *Journal of Documentation, 79*(1), 144-159. https://doi.org/10.1108/JD-06-2021-0118

Calder, B. J., Insko, C. A., & Yandell, B. (1974). The Relation of Cognitive and Memorial Processes to Persuasion in a Simulated Jury Trial. *Journal of Applied Social Psychology*, *4*(1), 62–93. https://doi.org/10.1111/j.1559-1816.1974.tb02600.x

Connor Desai, S., Xie, B., & Hayes, B. K. (2022). Getting to the source of the illusions of consensus. *Cognition, 223,* 105023. https://doi.org/10.1016/j.cognition.2022.105023

Eichler, A., Böcherer-Linder, K., & Vogel, M. (2020). Different Visualizations Cause Different Strategies When Dealing With Bayesian Situations. *Frontiers in Psychology*, *11*, 1897–1897. https://doi.org/10.3389/fpsyg.2020.01897

Epstein, Z., Pennycook, G., & Rand, D. (2020). Will the Crowd Game the Algorithm? Using Layperson Judgments to Combat Misinformation on Social Media by Downranking Distrusted Sources. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Cystems, 105,* 1-11. https://doi.org/10.1145/3313831.3376232

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. https://doi.org/10.1037/0022-3514.40.3.401

Harvey, Van Den Berg, D., Ellers, J., Kampen, R., Crowther, T. W., Roessingh, P., Verheggen, B., Nuijten, R. J., Post, E., Lewandowsky, S., Stirling, I., Balgopal, M., Amstrup, S. C., & Mann, M. E. (2018). Internet Blogs, Polar Bears, and Climate-Change Denial by Proxy. *Bioscience*, *68*(4), 281–287. https://doi.org/10.1093/biosci/bix133

Le Leu, H. (2021). *Examining the Impact of a Reasoning Aid to Help People Evaluate the Evidentiary Weight of Consensus*. (Undergraduate Honours Thesis). https://hdl.handle.net/2440/133226

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

Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and Its Correction: Continued Influence and Successful Debiasing. *Psychological Science in the Public Interest, 13*(3), 106-131. https://doi.org/10.1177/1529100612451018

Lombrozo, T. (2016). Explanatory Preferences Shape Learning and Inference. *Trends in Cognitive Sciences*, *20*(10), 748–759. https://doi.org/10.1016/j.tics.2016.08.001

Mercier, H., & Milton, H. (2019). Utilizing simple cues to informational dependency. *Evolution and Human Behavior, 40*(3), 301-314. https://doi.org/10.1016/j.evolhumbehav.2019.01.001

Nadarevic, L., Reber, R., Helmecke, A. J., & Köse, D. (2020). Perceived truth of statements and simulated social media postings: an experimental investigation of source credibility, repeated exposure, and presentation format. *Cognitive Research: Principles and Implications*, *5*(1), 56–56. https://doi.org/10.1186/s41235-020-00251-4

Pennycook, G., & Rand, D. G. (2019). Fighting misinformation on social media using crowdsourced judgments of news source quality. *Proceedings of the National Academy of Sciences, 116*(7), 2521-2526. https://doi.org/10.1073/pnas.1806781116

Pennycook, G., Bear, A., Collins, E. T., & Rand, D. G. (2020). The implied truth effect: attaching warnings to a subset of fake news headlines increases perceived accuracy of headlines without warnings. *Management Science*, *66*(11), 4944-4957. https://doi.org/10.1287/mnsc.2019.3478

Pew Research Center. (2022). *Social Media and News Fact Sheet.* Pew Research Center. https://www.pewresearch.org/journalism/fact-sheet/social-media-and-news-fact-sheet/

Pilditch, T. D., Hahn, U., Fenton, N., & Lagnado, D. (2020). Dependencies in evidential reports: The case for informational advantages. *Cognition*, *204*, Article 104343. https://doi.org/10.1016/j.cognition.2020.104343

Ransom, K. J., & Stephens, R. (2023). Tool Support for Reasoning about Consensus. *Proceedings of the Annual Conference of the Cognitive Science Society, 45*(45), 30-31.

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), 833-839. https://escholarship.org/uc/item/7x16q1r4

Shu, & 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. https://doi.org/10.1509/jm.11.0504

Simmonds, B. P., Stephens, R. G., Searston, S., 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), 828-834.

Vaupotič, N., Kienhues, D., & Jucks, R. (2021). Trust in Science and Scientists: Implications for (Higher) Education. In *Trust and Communication* (pp. 207–220). Springer International Publishing. https://doi.org/10.1007/978-3-030-72945-5_10

Weaver, K., Garcia, S. M., Schwarz, N., & Miller, D. T. (2007). Inferring the popularity of an opinion from its familiarity: A repetitive voice can sound like a chorus. *Journal of Personality and Social Psychology*, *92*(5), 821–833. https://doi.org/10.1037/0022-3514.92.5.821

Weisbuch, M., Mackie, D. M., & Garcia-Marques, T. (2003). Prior Source Exposure and Persuasion: Further Evidence for Misattributional Processes. *Personality and Social Psychology Bulletin*, *29*(6), 691–700. https://doi.org/10.1177/0146167203029006002

Xie, & Hayes, B. (2022). Sensitivity to Evidential Dependencies in Judgments Under Uncertainty. *Cognitive Science*, *46*(5), e13144. https://doi.org/10.1111/cogs.13144

Yousif, S. R., Aboody, R., & Keil, F. C. (2019). The Illusion of Consensus: A Failure to Distinguish Between True and False Consensus. *Psychological Science*, *30*(8), 1195-1204. https://doi.org/10.1177/095679761985684

Zemla, Sloman, S., Bechlivanidis, C., & Lagnado, D. A. (2017). Evaluating everyday explanations. *Psychonomic Bulletin & Review*, *24*(5), 1488–1500. https://doi.org/10.3758/s13423-017-1258-z

Appendix A

Pre-Registration

1) What's the main question being asked, or hypothesis being tested in this study?

Our aim is to investigate the assumptions people hold when reasoning from consensus. In particular, whether people demonstrate systematic preferences for the consensus quality cues of author frequency and argument frequency.

Two alternate arrangements or 'motifs' consisting of combinations of authors and reasons tweeting in response to a claim will be provided in a diagram. There will be six diagrams and each one will present two different motifs. In choosing a tweet to read from those available in a diagram, participants will be making a discrete choice between one motif or the other, thus indicating their preferences for the frequency of authors posting messages and the frequency of arguments made. The features of each diagram and a corresponding hypothesis of motif choice are discussed below. We will also explore the impact of tweet choice on agreement rating with the associated claim, measured before and after selecting a tweet from the diagram.

<u>Diagram 1:</u>
Motif 1- a single author giving different reasons. Motif 2- different authors each giving a different reason.

H1

- Participants will systematically prefer to read a message from a single author providing different reasons over different authors each providing a different reason.

<u>Diagram 2:</u>
Motif 1- a single author giving different reasons. Motif 3- different authors each giving the same reason.

H2

- Participants will systematically prefer to read a message from different authors each giving the same reason over a single author giving different reasons.

<u>Diagram 3:</u>
Motif 1- A single author giving different reasons. Motif 4- A single author repeating the same reason.

Literature suggests that people may infer that single authors who are able to provide different arguments (e.g., more expertise) are valuable. In contrast, people may also infer that one consistent argument that is repeated by an author is informative or useful. Given these conflicting positions, no hypothesis will be made for diagram 3.

<u>Diagram 4:</u>

Motif 2- Different authors each giving a different reason. Motif 3- different authors each giving the same reason.

H4

- Participants will systematically prefer to read a message from different authors each giving the same reason over different authors each giving a different reason.

Diagram 5:
Motif 2-. Different authors each giving a different reason. Motif 4-. A single author repeating the same reason.

H5

- Participants will systematically prefer to read a message from a single author repeating the same reason over different authors each giving a different reason.

Diagram 6:
Motif 3- different authors each giving the same reason. Motif 4- A single author repeating the same reason.

H6

- Participants will systematically prefer to read a message from different authors each giving the same reason over a single author repeating the same reason.

2) Describe the key dependent variable(s) specifying how they will be measured.

There will be two dependent measures in the study. The first will be participants' choice of tweet that they most want to read after viewing the diagram. This measure will be used to record participants' preferences for author and reason frequency in a consensus. The second dependent measure will be participants' agreement with a claim on a scale from -50 ("strongly disagree") to 50 ("strongly agree"). This will be measured before and after viewing the diagram and making a tweet choice. This measure will be used to collect initial agreement ratings with a claim and to observe the impact of tweet selection on agreement ratings.

3) How many and which conditions will participants be assigned to?

In a Discrete Choice Experiment design, two factors of author frequency (same or different authors) and reason frequency (same or different reasons) will be crossed to form 4 motifs, which are then presented in diagrams showing the 6 possible pairwise combinations, manipulated within-subjects and between-trials. Participants will see each diagram three times, forming 18 trials.

4) Specify exactly which analyses you will conduct to examine the main question/hypothesis.

We will use binomial tests (test of equal proportions) with 95% confidence intervals to test whether participants as a group have consistent preferences for the frequency of authors and arguments in a consensus. We will test whether participants' observed motif choice in each of the six diagrams deviates from what would be expected by chance, using a different test for each of the six diagrams and hypotheses.

Any analysis that will be performed on updated agreement ratings based on motif choice, dependent upon observed choice variation, will be exploratory in nature.

5) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.

Participants who do not complete all 18 trials will be excluded.

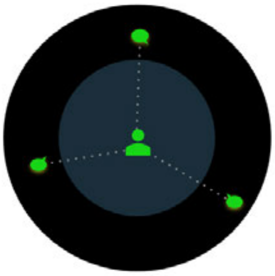
6) How many observations will be collected or what will determine sample size?

We will recruit 100 adult participants using the Amazon Mechanical Turk platform.

Appendix B

Figure B1.

The Six Multiple-Choice Questions Presented to Participants After Experiment Instructions.



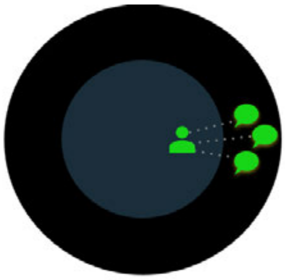
Please select the scenario displayed above:

a) One author giving the same reason

b) One author giving different reasons

c) Three authors giving the same reason

d) Three authors giving different reasons



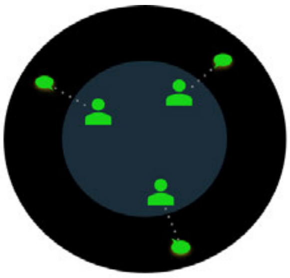
Please select the scenario displayed above:

a) One author giving the same reason

b) One author giving different reasons

c) Three authors giving the same reason

d) Three authors giving different reasons



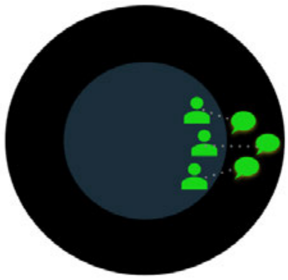
Please select the scenario displayed above:

a) One author giving the same reason

b) One author giving different reasons

c) Three authors giving the same reason

d) Three authors giving different reasons



Please select the scenario displayed above:

a) One author giving the same reason

b) One author giving different reasons

c) Three authors giving the same reason

d) Three authors giving different reasons

Regarding the claims we ask you to evaluate…?

a) We require an answer accurate to 2 decimal places

b) We have specific answers in mind that you need to get correct

c) We don't have any right or wrong answers in mind

Regarding the sliding scale that we use…?

a) -50 means "Strongly disagree" and 50 means "Strongly agree"

b) -50 means "I'm not confident in my answer and 50 means "I'm sure I got the answer right"

c) -50 means "I agree" and 50 means "I disagree"

d) This is a trick question. Sliding scales aren't used.

Appendix C

Table C1.

Independent Samples t-tests for Differences in Mean Agreement Change Between Motif Choices.

Diagram	Motifs	Mean	SD	t	df	Significance	Effect Size *(d)*
1	Motif 1	6.93	12.09	1.01	288	0.31	0.11
	Motif 2	5.61	10.38				
2	Motif 1	6.91	12.91	0.30	172	0.76	0.03
	Motif 3	6.43	12.51				
3	Motif 1	4.19	10.29	-0.07	255	0.93	-0.009
	Motif 4	4.30	12.18				
4	Motif 2	6.35	12.22	0.82	124	0.41	0.11
	Motif 3	5.00	12.16				
5	Motif 2	3.81	10.12	-1.01	294	0.30	-0.11
	Motif 4	5.04	10.69				
6	Motif 3	4.98	10.45	-1.27	132	0.20	-0.17
	Motif 4	6.96	12.88				

Figure C1.

Difference (Delta) in Mean Agreement Change Between Motif Choices for the Six Diagrams.

Diagram 1

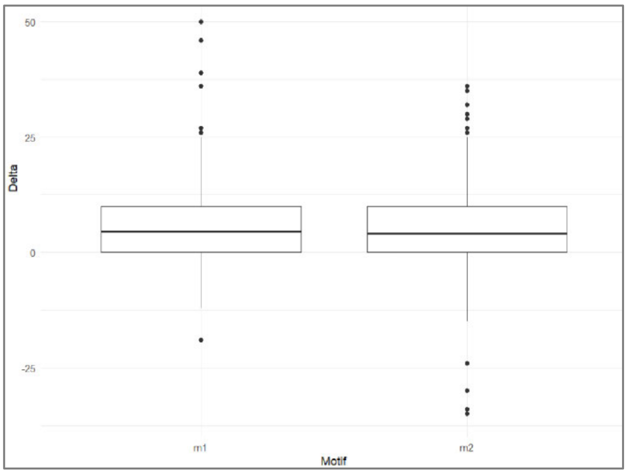


Diagram 2

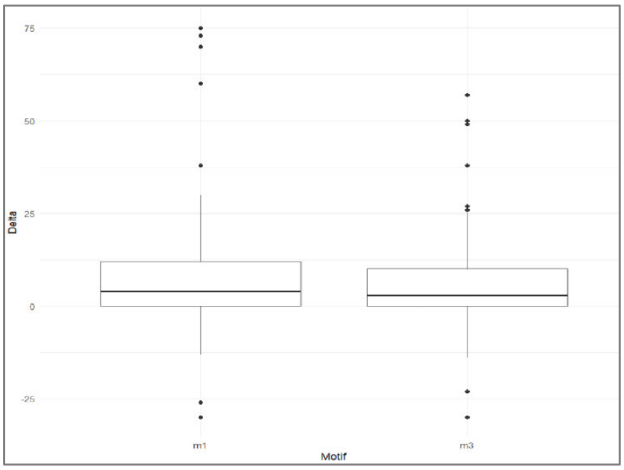


Diagram 3

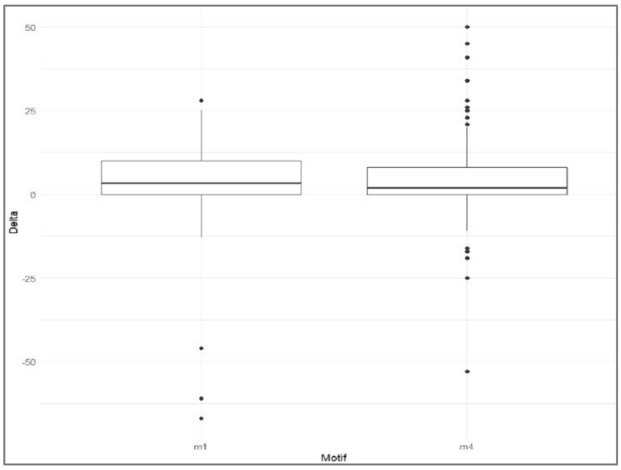


Diagram 4

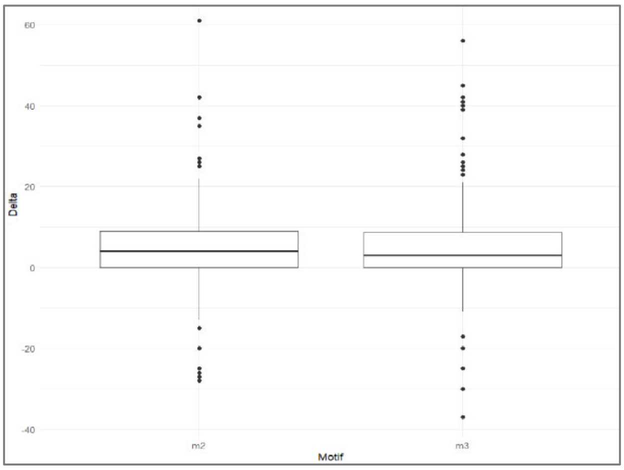


Diagram 5

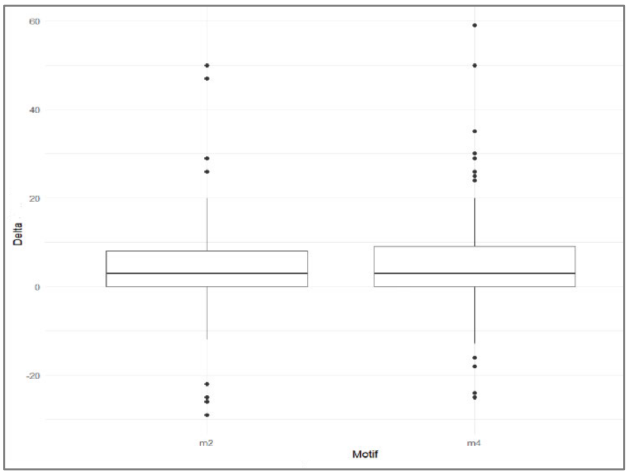


Diagram 6