How convincing is a crowd? Quantifying the persuasiveness of a consensus for different individuals and types of claims

Manikya Alister[1], Keith Ransom[4], Saoirse Connor Desai[2], Ee Von Soh[1], Brett Hayes[3], and Andrew Perfors[1]

[1]The University of Melbourne

[2]The University of Technology Sydney

[3]The University of New South Wales

[4]The University of Adelaide

Word Count: 2,000 (introduction + discussion)

Author Note

Abstract

A powerful cue when reasoning is whether an apparent consensus has been reached. However, we do not yet know how the strength of this cue varies between different individuals and types of claims. The current study (n = 78 US adults recruited from Prolific) evaluated this with a realistic mock social media paradigm where each participant evaluated 60 diverse, real world claims based on posts from people that either disagreed with each other, formed a consensus independently, or formed a consensus using shared sources. Almost all participants revised their beliefs in line with the consensus, with many also qualitatively changing their minds. A consensus was also more persuasive for claims more likely to have a ground truth (more "knowable" claims). While most people were insensitive to consensus independence, some were more persuaded by a consensus formed independently whereas some were equally as convinced by a consensus formed using the same sources.

Keywords: consensus; persuasion; source independence; social reasoning; individual differences

Researcher transparency statement

General Disclosures

Conflict of interest. All authors declare no conflicts of interest.

Funding. Manikya Alister was supported by an Australian Government Research Training Program Scholarship.

Artificial Intelligence. Artificial Intelligence (chatGPT; more details provided in the Procedure) was used to generate some of the fake social media posts used in the experiment. However, the originally generated posts were later refined manually to ensure realism and variety. No other artificial intelligence assisted technologies were used in this research or the creation of this article.

Ethics. Our study was approved by the University of Adelaide Ethics Committee (Approval No. 20/78).

Open Science Framework (OSF). To ensure long-term preservation, all OSF files were registered at `osf.io/f92ja`

Study Disclosures

Preregistration. Our research aims, experimental design and most of our analyses were preregistered on the 8th of January, 2024. Primary data collection began on the 10th of January. All preregistered analyses were performed as described with no deviations, and analyses that were not preregistered are clearly specified when mentioned in the main manuscript. Our preregistration can be found here: `aspredicted.org/7pdm-vgpx.pdf` . As noted in our preregistration, we collected pilot data from 5 participants to ensure that the experiment was running correctly and our estimates about how long it takes were correct. We did not make any changes to the experiment after piloting, so these participants were included in the final sample.

Materials. All of the experimental material (social media posts, university/media institutions, user names, and photos) are publicly available at `osf.io/mtuyv/`

Data. All anonymized primary and clean (including cleaning scripts) data are available at `osf.io/mtuyv/`.

Analysis Scripts. All analysis scripts necessary to reproduce the results are available at `osf.io/mtuyv/`.

Computational reproducibility. The computational reproducibility of the results has been independently confirmed by the journal's STAR team.

How convincing is a crowd? Quantifying the persuasiveness of a consensus for

different individuals and types of claims

When encountering a new opinion or claim, it is often difficult to quickly and accurately verify its truth based on personal experience alone. We might approach this situation by relying on cues such as the number of people who agree with it. For instance, if you come across a social media post stating that 'genetically modified crops are a good idea,' you might be inclined to give this claim more weight if there is a consensus supporting it.

This kind of "consensus" effect, where people tend to be more convinced by something that multiple people agree with, has been demonstrated extensively (see Mercier & Morin, 2019; Oktar & Lombrozo, 2025, for reviews). However, reasoning about consensus in everyday life can be complex. One such complexity is that people vary in how much they engage in critical reasoning generally (Cacioppo & Petty, 1982; Kruglanski et al., 1993). Another complexity is that *many* factors go into evaluating complex claims, and individuals differ in how they weigh those factors. These factors include the prestige (Atkisson et al., 2012) of or confidence (Sah et al., 2013) in the source of the claim, and the complexity of the arguments given in support of it (Zemla et al., 2017). Moreover, reasoners realize that people making arguments in support of a claim (e.g., on social media) have both different levels of competence (Lin et al., 2016), *and* goals, which include persuasion, identity signaling, trolling, and deception (Pucci et al., 2023). If individuals have different assumptions about the nature or importance of these factors, this will mean that they differ in how sensitive they are to consensus effects in the first place. Indeed, research suggests people differ in the extent to which they consider social information when making judgments and decisions (Molleman et al., 2019; Toelch et al., 2014). However, previous research has typically focused on group-level rather than individual behavior.

Another key factor is the independence between the sources that contribute to the

consensus. It is often argued that a consensus should be more convincing when everyone within the consensus reached their opinions *independently* from each other (Connor Desai et al., 2022; Harkins & Petty, 1987; Whalen et al., 2018; Xie & Hayes, 2022; Yousif et al., 2019). If multiple social media posts from different people agree that genetically modified crops are a good idea, and they all reference different, independent sources supporting this claim (e.g., different scientific studies or surveys), one would think that this sort of "independent consensus" *should* be more convincing than if each of them referenced the same source ("dependent consensus"). Indeed, normative models of decision-making support the idea that people should weigh claims corroborated by multiple independent sources more than dependent (repeated) sources (for a review, see Couch, 2022).

Unfortunately, in the real world, it is often unclear which primary sources have influenced someone's opinions. Even if the primary sources that influenced people are known, one might be unsure whether those sources are truly independent, since they could have collaborated or used the same underlying data (Madsen et al., 2020; Pilditch et al., 2020). These difficulties mean that people might believe that a consensus is not independent when it actually is, or that what looks like an independent consensus actually is not. For instance, most COVID-19 anti-vaccination views originated from the same few people (Center for Countering Digital Hate, 2021) and the majority of climate denial blogs rely on the same few primary sources (Harvey et al., 2018).

Given these implications, it is vital to understand the extent to which people are actually sensitive to consensus independence when reasoning about real world topics. Indeed, this question has been the focus of several recent experimental investigations. Much of this evidence suggests that people are *insensitive* to consensus independence except in specific contexts, such as when the claim is easily knowable (Aboody et al., 2022; Yousif et al., 2019), independence is clearly emphasized (Connor Desai et al., 2022), or when the relationship between the source and the conclusion is clear (Alister

et al., 2022) — but even then, the effect of source independence is usually very small.

However, the issue is still far from settled. Firstly, since it is usually assumed that an independent consensus *should* be the more persuasive form of evidence, relatively little attempt is usually made to ascertain the extent or nature of individual variation from this standard. Indeed, pure repetition of information, regardless of source independence, can also influence beliefs (e.g., Fazio et al., 2022; Pillai & Fazio, 2024). As discussed above, people might reasonably differ in how they use source independence to judge reliability or competence. This individual variation may look like a null or weak effect on the group level but reflect interesting and sophisticated reasoning at the individual level (Xie & Hayes, 2022).

A second limitation is that studies showing the "standard" consensus effect (i.e., stronger belief in claims that are endorsed by many people) have typically involved a limited number of claims and claim types. This issue is important because people might reason about consensus differently for different kinds of claims (Richardson & Keil, 2022). For instance, Yousif et al. (2019) found that when a claim was about a new tax policy and the sources were economists, people were more sensitive to dependence than when it was about an event at a local school and the sources were eyewitnesses. The authors concluded that people may reason differently about claims that are less knowable (like an economic prediction) than those that have a clear ground truth (like eyewitness accounts). This hypothesis is consistent with work demonstrating that the source expertise matters (Maddux & Rogers, 1980; Simmonds et al., 2024), probably in part because experts have more insight into the ground truth of a situation. However, claim knowability has never been systematically manipulated over a variety of claims, nor has this hypothesis been tested in a more standard consensus paradigm comparing contexts where a consensus about a claim has or has not been reached. Hence, it remains unclear how much claim knowability matters or how robust this effect is.

The current study addresses these limitations and is the first to systematically

investigate individual and claim level differences in consensus effects for a large number of realistic claims. The current study involved, to our knowledge, by far the most stimuli in any experimental consensus reasoning paradigm to date. Each of our participants read a total of 420 unique, realistic social media posts, out of a total pool of 1,080 total posts (to allow for appropriate randomization across conditions) for 60 unique claims (one claim per trial; see Method). Generating this many trials enabled us to implement reliable Bayesian analyses for each person. Including more claims also allowed us to develop and test a new taxonomy of claim types based on their "knowability", which meant we were able to systematically identify whether different kinds of claims elicited different consensus effects.

Method

Our study was approved by the University of Adelaide Ethics Committee (Approval No. 20/78).

Participants

160 participants were recruited from Prolific Academic. Of those participants, 116 completed session 1 and were paid £5.25 per session for up to two 35 minute sessions. 38 participants were removed based on our preregistered exclusion criteria. 25 participants were removed due to having lower than 90% accuracy in our comprehension checks (see Procedure) in session 1, and were not invited to complete session 2 (except for 2 pilot participants who did both sessions despite having low accuracy in the first session, but were still removed from analyses). Of the remaining participants who were invited to complete Session 2, a further 2 participants were removed due to low accuracy and 11 were removed due to not completing Session 2, leaving a total of 78 participants. Ages ranged from 18 to 72 years old ($M = 36$) with 42 participants identifying as female, 34 as male, and 2 otherwise. All were pre-screened as being both native English speakers and English being their primary language.
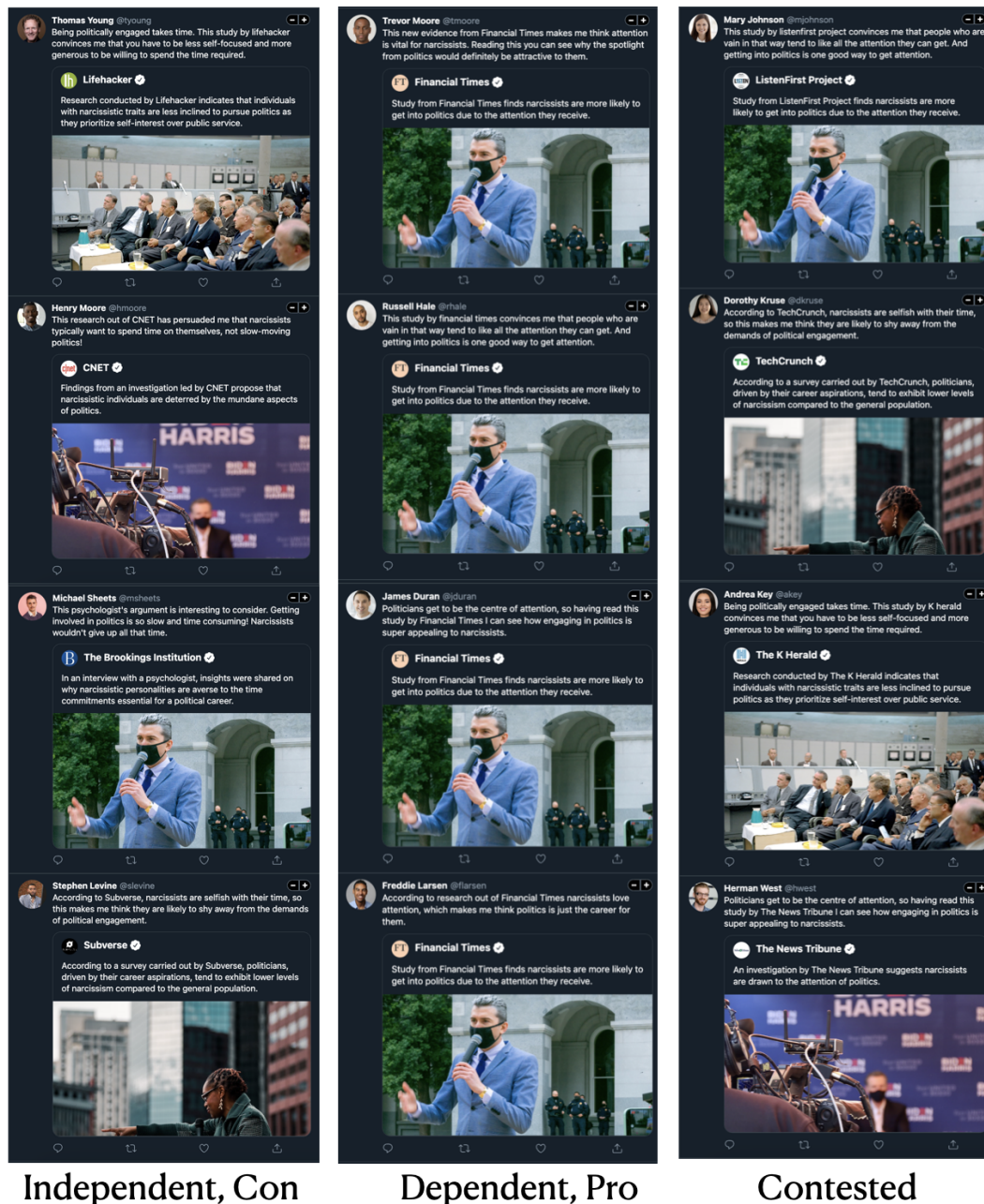
Claim: Narcissists are more politically engaged

Independent, Con   Dependent, Pro   Contested

Figure 1

Experiment stimuli. *Sample posts from each consensus condition. The left panel is from the* INDEPENDENT *consensus condition, where each person re-posted a different source (here, arguing in against the claim). Those in the middle are from the* DEPENDENT *consensus condition, where each person re-posted the same source (here, arguing for it). The right column shows the* CONTESTED *condition, where an equal number of people support/oppose the claim. The sources, names, photos, text, dependence, and support direction were randomized for each person and claim. In this example, the sources are news organizations, but for some claims they were universities.*

Procedure

After providing consent and passing a short quiz regarding the task instructions, each participant saw 60 trials over the course of two sessions on two separate days (30 trials per day). Our choice in the number of trials was informed by a simulation-based power analysis (see Supplemental Material 2). Each trial began with participants viewing a claim (e.g.,"Narcissists are more politically engaged"). The claim was accompanied by a photo and a brief, neutral social media post pertaining to the claim (e.g., "I'm a modest person, many people say I'm incredibly modest, so I found this interesting to consider...") to help set the scene for the participant and facilitate engagement with the claim. After reading the claim, they were asked to indicate "to what extent do you agree with the claim?" using a slider from 0 to 100. They then viewed four social media posts by four distinct users who were each sharing another post from a primary source (see Figure 1). To ensure that people were engaging with the task properly, they had to indicate whether each user was arguing for or against the claim. The 27 participants who were less than 90% accurate at this task were removed from the analysis.

As shown in Figure 1, each post took the form of a re-post and included the primary source being re-posted, the primary data referred to by the source, and the user's own words explaining how the primary source persuaded them. Because the four users were always distinct people with unique profile photos and names, this made it clear that the user had read the source and that it had influenced their opinion about the claim; this was shown to be important in previous work (Anonymous, Year).

After confirming they had read each post (by indicating whether the post endorsed or opposed the claim), participants were once again asked to indicate how much they agreed with the claim (using the same 0-100 slider as before, but initialized to indicate their initial rating). The difference between a participant's rating before and after seeing the claims represented their degree of belief revision due to the posts. After completing all 60 trials of the experiment, they were given a multiple-choice question

about which strategy they used to evaluate the claims (see Results for full details of the question and response options).

Consensus conditions

Our primary manipulation (within participants) pertained to the presence or absence of a consensus within the four social media posts as well as – if there was a consensus – whether or not it was formed independently. Specifically, trials were either consensus trials, where everyone reached the same opinion, or CONTESTED trials, where an equal number of posts supported and opposed the claim (right column of Figure 1). There were two kinds of consensus trials: INDEPENDENT consensus and DEPENDENT consensus (20 trials each per person). In INDEPENDENT trials, each of the four social media posts cited different primary sources (left column of Figure 1). In DEPENDENT trials, each of the four social media posts cited the same primary source (middle column of Figure 1). The stance of the consensus was randomized, with the constraint that there was always an equal number of supporting (PRO) and opposing (CON) trials for each claim type and consensus condition.

The 20 CONTESTED trials were included for two reasons. First, they allowed us to investigate individual differences in the strength of standard consensus effects (i.e., how many people showed a larger belief shift in full consensus vs no consensus trials). Second, it served to reduce demand effects and ensure participants read all of the posts; otherwise they could pass the manipulation check without reading them once they realized that all posts already agreed with the first one.

Regardless of condition, all four posters and sources gave essentially the same reason in different words (e.g., those arguing PRO on the narcissism claim all pointed out that narcissists were more likely to get into politics due to the attention they receive, and those arguing CON all pointed out that the amount of time required to engage in politics is unappealing for narcissists; see Figure 1). The order of claims and posts, assignment of claims to condition, and assignment of avatars and names was randomized across all

participants and claims. The primary source was always either a news organization or a university, and the primary data was always some kind of study or investigation carried out specifically by that organization (made clear through the wording of the post). For example, the primary data could be a study by the University of Springfield, and the primary source would be the official account of the University of Springfield.

In the DEPENDENT condition, each of the four posts (by four different users) re-posted the same article by the same source (hence the same primary data). In the INDEPENDENT and CONTESTED conditions, the source of the post that was re-posted and the data that the source referred to were both distinct for each of the four posts: Person A cited Source X, Person B cited Source Y, and so forth. Thus, each was re-posting an independent source and referring to independent primary data. There were always three primary sources and one expert testimony, which was included to add some variety and reduce demand effects. We tried to maintain a balance between how many trials used each source type, but in some cases it only made sense to have a particular source type (e.g., a university would not conduct a study about whether a mayor ran into a burning building). In total, 27 of the claims had universities as their primary sources, and 33 had news organizations.

The news companies were real media companies chosen via the website AllSides,[1] which allows people to rate the bias of different news companies. We chose news companies that were mid-range in popularity and deemed "centrist" by the raters. The universities were a sample of real universities ranked between 100 and 200 by the QS World University Rankings. These universities were chosen so that they would be recognizable enough such that they were believable, but not so much so that their reputation would substantially influence people's beliefs. Although whether the sources were news companies or institutions on each trial was deliberately chosen based on the appropriateness to the claim, the actual company/institution included in each trial was

---

[1] allsides.com/unbiased-balanced-news

randomized for each participant and were always unique. The companies/institutions always had a profile photo, full name, and a "verified" tick to signal authenticity. All content was fictitious. The content that was not part of Anonymous (Year) was first generated using chatGPT-3.5-turbo-1106 and then refined manually to ensure realism and variety.

Claim type conditions

In addition to varying the nature of the consensus across conditions within-participant, we also varied the nature of the claims themselves. Given that previous research suggests knowable claims are more likely to induce consensus independence effects (Yousif et al., 2019), we selected 30 *knowable* claims and 30 *unknowable* ones. Each category was further subdivided based on the way in which it was knowable or not. For instance, KNOWABLE EYEWITNESS claims include something that somebody could have seen (e.g., a mayor saving a child from a burning building). A KNOWABLE FACT is something that while verifiable in principle, you would need to be an expert in order to do so (e.g., a new species of jellyfish being discovered). An UNKNOWABLE EXPERT claim is one which does not have a known ground truth at the moment, but nevertheless expertise would be helpful in evaluating (e.g., economic forecasting). Lastly, UNKNOWABLE PREFERENCE claims do not have a ground truth and expertise is less likely to be important (e.g., whether flying is a better super power than invisibility). Thus, each of the 60 total trials per participant were comprised of 15 trials of each of the four claim types. Given previous literature (Yousif et al., 2019), we expected that knowable claims would result in larger consensus effects compared to unknowable claims. The range and strength of prior beliefs endorsed by participants varied considerably across claims and within claim types. (see Supplemental Material 6 for the full set of claims and the distribution of people's initial belief ratings for them).

Results

Aggregate behavior

We first examined the extent to which belief revision was affected by the knowability of a claim and the type of consensus. To account for the fact that CON trials (where the consensus argued against the claim) would shift beliefs in the opposite direction, for the purpose of analysis the belief scores for CON trials were reversed so that a positive shift always means an update in the direction of the consensus. Effect sizes were equivalent for PRO and CON trials across conditions (see Supplemental Material 7.1).

As Figure 2 shows, the largest belief updates occurred in the full consensus trials (INDEPENDENT and DEPENDENT, where all four posts took the same stance towards the claim). People were much more persuaded by these trials than CONTESTED ones in which half of the posts argued in one direction and half in the other. In the CONTESTED trials, people actually tended to believe the claim *less* after seeing the posts, although the degree of belief revision was very small. The figure also suggests the presence of a small but consistent difference between INDEPENDENT and DEPENDENT consensus trials, with participants on average more convinced when sources are independent.

To quantitatively assess the persuasiveness of different kinds of consensus as a function of the type of claim, we compared four nested Bayesian generalized linear models using the brms (2.20.4; Bürkner, 2018) package in R (4.2.2), using default priors: a weakly informed student's t-distribution for the intercept and a uniform (flat) prior for the coefficients (slopes). Inspection of convergence metrics, such as $\hat{R}$, suggested that using flat priors did not cause any problems with convergence in the sampling process.

The outcome variable was the rating after reading the four posts and people's initial beliefs (before seeing the posts) were included as predictors in all models (e.g., Table 2). To assess the relative performance of each model, we compared them using

the leave-one-out cross-validation criterion (LOOIC; Vehtari et al., 2017). Our focus was

on two kinds of consensus effects: Standard consensus (INDEPENDENT trials vs.

CONTESTED trials); and Consensus independence (INDEPENDENT trials vs. DEPENDENT

trials). We therefore ran each of the models on the relevant subset of data, as described

in the next section. Although we did not conduct a simulated power estimation for the

analyses of the aggregate behavior like we did for the individual-level analyses (see

Supplementary Material 2), we expected this sample size to be sufficient for calculating

group level effects given the number of trials per person, since previous studies using the

same design have been able to detect similar sized effects with around 1/4 of the

number of total observations in our study (e.g., Anonymous, Year).

　　　As we have mentioned, we chose a Bayesian statistical approach (both here and

in the individual-level analyses presented later). We chose this instead of a frequentist

approach for several reasons. First, Bayesian models estimate parameters by calculating

the posterior probability distribution for each parameter, which gives the probability of

different parameter values given the data and our prior expectations (for an overview of

Bayesian analysis, see McElreath, 2016). This contrasts with the frequentist approach,

which focuses on testing hypotheses and provides p-values, or the probability of

observing data as extreme as the observed data under the null hypothesis. Frequentist

methods, such as null hypothesis significance testing (NHST), only allow us to reject or

fail to reject the null hypothesis, often with the sole conclusion of whether a parameter is

different from zero (e.g., $p < 0.05$).

　　　In contrast, Bayesian methods allow us to quantify the uncertainty surrounding

the entire range of possible parameter values, not just whether a parameter is

significantly different from zero. Bayesian analyses yield credible intervals—the range of

parameter values within which the true value lies with a specific probability, given the

data and the model. For example, the 89% credible intervals we report in Tables 2, 4,

and Figure 4 specify that there is an 89% probability that the true parameter lies within

this range, given the observed data and prior assumptions (we provide an explanation for why "89%" in Supplementary Material 2).

This interpretation is more intuitive than the frequentist 95% confidence interval, which is often misunderstood. A confidence interval reflects the range within which the parameter would fall if we could repeatedly sample from the population and calculate estimates in each sample. Importantly, it does not provide direct probabilistic information about the parameter given the data at hand, which is what Bayesian credible intervals *do* give us.

Comparing beliefs in full consensus vs no consensus

We first asked whether people changed their belief more when the four posts agreed (a full consensus) than when half argued in one direction and half in the other (no consensus). In order to quantitatively test this, we compared the INDEPENDENT condition trials to the CONTESTED condition trials which were directly comparable in other ways: in both conditions, the four posts had distinct, independent sources. (DEPENDENT trials differed from CONTESTED trials by having a single repeated source and were thus excluded from this comparison.)

Results of the full model comparison are in Table 1. The best model (M4), whose coefficients are reported in Table 2, had credible main effects of both consensus type (INDEPENDENT vs CONTESTED) as well as an interaction with claim type. The main effect of consensus suggests that people were more convinced by INDEPENDENT trials compared to CONTESTED trials. The main effect of claim type suggests people were most convinced by KNOWABLE Eyewitness claims, followed by KNOWABLE Facts, then UNKNOWABLE Expert, and lastly UNKNOWABLE Preference. Follow up comparisons, reported in Supplemental Material 3, revealed credible differences between each claim type.

The model also suggested an interaction between the nature of the consensus and the type of claim: the difference in belief between a full consensus and no
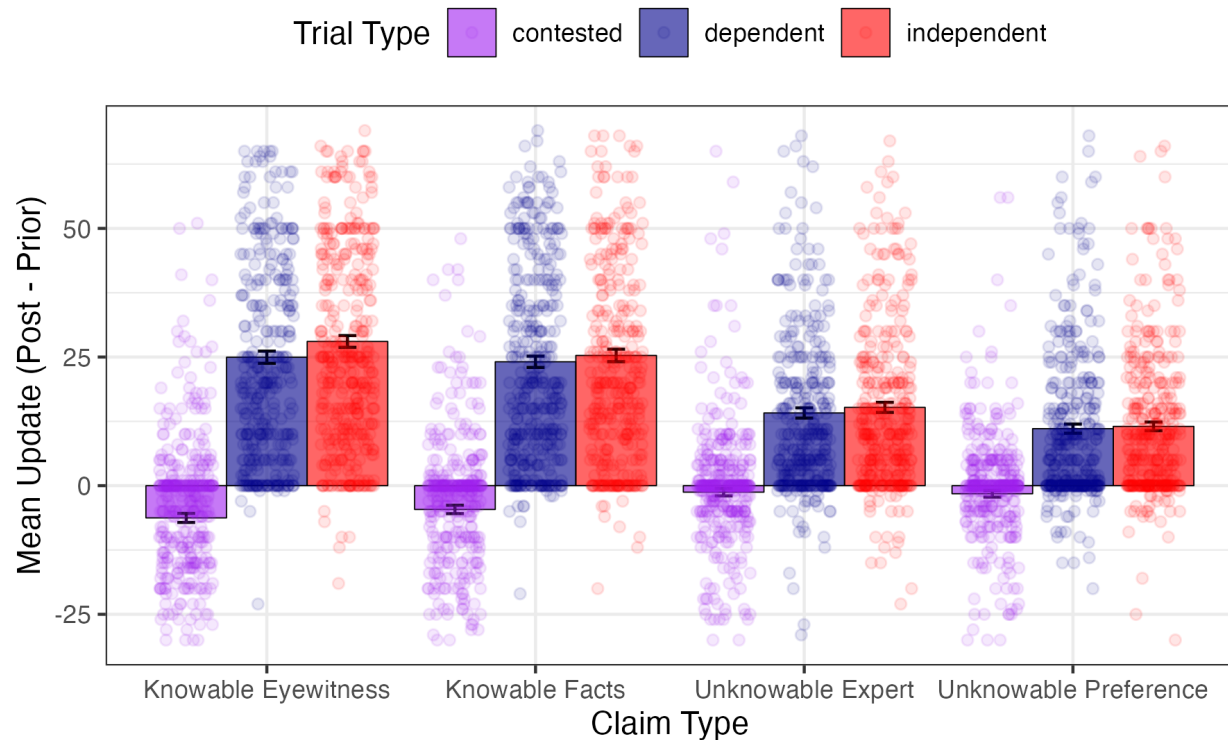
Figure 2

Belief update as a function of trial type and claim type. *Positive belief update indicates belief change in the direction of the consensus for the full consensus trials.* INDEPENDENT *trials (red) showed four users who shared the same opinion but cited different primary sources.* DEPENDENT *trials (blue) showed four users who shared the same opinion and cited the same primary source.* CONTESTED *trials (purple) showed four users, half agreeing with the claim and half disagreeing, with each user citing different sources. The axes have been slightly constrained to better show the mean differences, so some individual data points are not visible.*

consensus was larger when the claims were more knowable. Interestingly, looking at Figure 2, this interaction reflects not only the fact that for knowable claims people were more convinced by a consensus, but also that they were more likely to revise their beliefs *against* knowable claims on contested trials. One might imagine that this reflects initial differences in certainty between knowable and unknowable claims. However, this does

Table 1

Model comparison predicting belief in the claim after INDEPENDENT ***versus*** CONTESTED

trials. *M1 only considered participants' initial beliefs about the claim. M2 also considered*

whether there was a consensus or not, and M3 added claim type. M4 also considered the

interaction between the presence of a consensus and claim type and was favored by LOOIC. All

models also contained a random intercept term for each participant (Brown, 2021).

Model	LOOIC	SE	Rank
M1. InitialBelief	27693	100	4
M2. InitialBelief + Consensus	26633	116	3
M3. InitialBelief + Consensus + ClaimType	26565	117	2
M4. InitialBelief + Consensus $\times$ ClaimType	**26335**	121	1

not appear to be the case: when we added prior certainty to the models (operationalized

as the distance of initial beliefs to the initial belief scale midpoint, (as in Orticio et al.,

2022) the main effects and interactions remained (see Supplemental Material 4.3).

Instead, what might be happening is that the presence (or absence) of a consensus is

more informative for knowable claims. If a claim is unknowable in principle, it doesn't

mean much whether people agree or disagree. Conversely, full consensus for a

knowable claim might be reasonable evidence that the claim is true, while disagreement

on knowable claims is a good indication that it might not be.

Comparing beliefs in dependent vs independent consensus

Our second question was whether people reasoned differently for INDEPENDENT

and DEPENDENT trials. As shown in Table 3, the full model comparison favored model

M3, indicating that the independence of the consensus and the type of claim were both

important to belief revision (with no interaction): on average, people tended to be more

convinced by an INDEPENDENT consensus than a DEPENDENT one. The main effect of

claim type reflected a larger effect for knowable than unknowable claims (Figure 2).

Although the main effect of independence was credibly greater than zero in the winning

model, it was quite small relative to the other consensus comparison (as evident in

Figure 2 as well as the coefficients reported on page 2 of Supplemental Material 3).

A follow-up analysis indicated that the group level results were qualitatively

unchanged even on only the first trial, suggesting that they did not arise due to demand

effects or participants becoming sensitized to the consensus manipulations over the

course of the experiment (the full analysis is reported in Supplementary Material 7.2).

Table 2

Unstandardized coefficients for the best model comparing INDEPENDENT *to* CONTESTED

Coefficient	Estimate	89% Credible Interval	
		Lower	Upper
Intercept	9.54	7.82	11.26
InitialBelief	0.72	0.70	0.73
Consensus_{Ind}	32.90	31.12	34.68
$\text{ClaimType}_{KnowFact}$	1.91	0.03	3.76
$\text{ClaimType}_{UnknowExpert}$	6.40	4.59	8.21
$\text{ClaimType}_{UnknowPref}$	4.21	2.33	6.02
$\text{Consensus}_{Ind} \times \text{ClaimType}_{KnowFact}$	-4.65	-7.20	-2.06
$\text{Consensus}_{Ind} \times \text{ClaimType}_{UnknowExpert}$	-18.78	-21.36	-16.17
$\text{Consensus}_{Ind} \times \text{ClaimType}_{UnknowPref}$	-21.49	-24.13	-18.80

Note. Because InitialBelief is a continuous variable on the same scale as the outcome measure, it is interpreted as: every unit increase in people's initial beliefs was associated with a .72 increase in their beliefs after seeing the posts (therefore suggesting that participants' initial beliefs accounted for quite a lot). The coefficients of the categorical variables are interpreted with respect to their *reference* variables, which is the level of the variable not contained in the table. Further, because of the default dummy coding of the interaction model, the main effects are interpreted at the interacting variable's reference level, rather than aggregating across the other variable. So, 32.9 for Consensus_{Ind} means that participants rated INDEPENDENT trials as 32.9 units more convincing than CONTESTED trials when the claim type was knowable eyewitness. For the same reason, the claim type coefficients reported here correspond to their values when the consensus trial is CONTESTED, which is why the estimates suggest that KNOWABLE EXPERT trials were the least convincing, even though when aggregating across all independence conditions they were the *most* convincing (see last page of Supplemental Material 3 for full post hoc comparisons). Following Kruschke (2014), we show 89% credible intervals in all analyses (see simulated power estimation in Supplementary Material 2 for further justification).

Table 3

***Model comparison predicting belief in the claim after* INDEPENDENT *versus* DEPENDENT**

trials. *The models were similar to Table 1, but compared the two Independence conditions*

instead. M3 was favoured by LOOIC.

Model	LOOIC	SE	Rank
M1. InitialBelief	26952	103	4
M2. InitialBelief + Independence	26946	103	3
M3. InitialBelief + Independence + ClaimType	**26493**	111	1
M4. InitialBelief + Independence $\times$ ClaimType	26496	111	2

Individual Differences

A key aim of this experiment was to explore whether and how individuals vary in how they reason about consensus. We therefore looked at individual behavior on the same questions as before: To what extent do people change their beliefs more when there is a full consensus vs none? Or an INDEPENDENT vs DEPENDENT one? As Figure 3 shows, the vast majority of people were sensitive to consensus, changing their belief more when all posts agreed; however, people varied substantially in how they reasoned about independence.

We used Bayesian linear models to quantitatively classify the different kinds of participants as one of four possible logical types. First, some people might be generally insensitive to consensus: they reason similarly regardless of whether all four people agree or there are two on each side. Second, some people might be sensitive to the presence of a consensus but insensitive to source independence. Third, some people might be more convinced by a consensus when the sources are INDEPENDENT. And finally, some might be more convinced when the sources are DEPENDENT. Each participant was fit to a model in which the outcome variable was that person's belief after

Table 4

***Unstandardised coefficients for the best model comparing* DEPENDENT *to* INDEPENDENT**

Coefficient	Estimate	89% Credible Interval	
		Lower	Upper
Intercept	42.896	40.789	44.977
InitialBelief	0.666	0.648	0.684
Consensus$_{Ind}$	1.678	0.725	2.620
ClaimType$_{KnowFact}$	-2.706	-4.055	-1.329
ClaimType$_{UnknowExpert}$	-12.420	-13.816	-11.054
ClaimType$_{UnknowPref}$	-15.898	-17.271	-14.512

A detailed note on how to interpret the coefficients is in Table 2. The only difference is that since the best-performing model did not contain an interaction, the main effects of the categorical variables are collapsed across the levels of the other variable.

seeing the posts. We then compared the two models: a baseline Null model in which the only predictor was that person's prior belief in the claim, compared to an Alternative model that also included a predictor corresponding to the effect in question.

The top panel of Figure 4 explores sensitivity to consensus (whether people changed their beliefs more with a full consensus than none). It shows, for each person, whether they were best fit by the Null model (blue circle) or by the Alternative (green triangle) that included a predictor corresponding to Consensus (CONTESTED vs INDEPENDENT). All but three participants (96%) were best described by the Alternative model, suggesting that the vast majority of people were more convinced by a four people agreeing than by two people on each side. The median individual difference in belief change between the full consensus (INDEPENDENT) and CONTESTED trials was 23 points in favor of the INDEPENDENT trials. There was substantial individual variation,

however, ranging from close to zero change to over 40 points of shift.

The bottom panel of Figure 4 explores sensitivity to independence. It shows, for each person, whether they were best fit by the Null or an Alternative model that included a predictor corresponding to Independence (DEPENDENT vs INDEPENDENT). The majority of people were best fit by the Null model, but 22% were more persuaded on INDEPENDENT trials (positive on the y axis) and 8% were more persuaded on
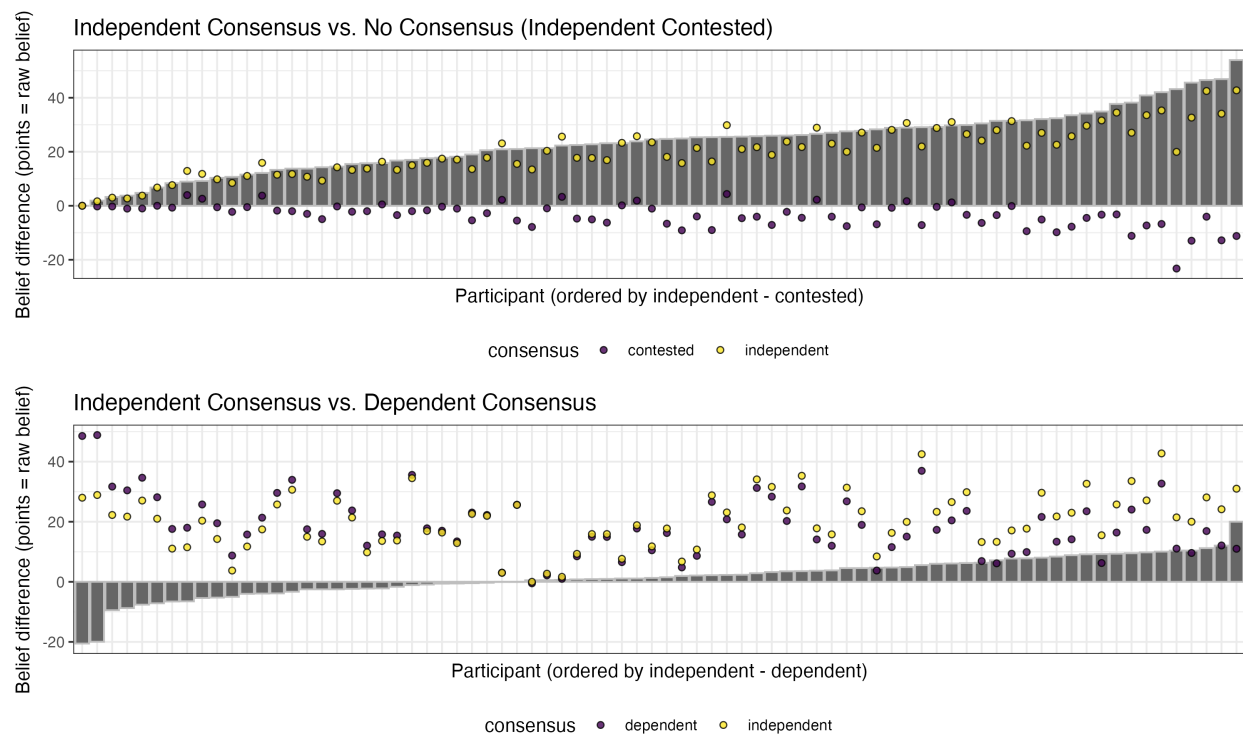

Figure 3

Belief updates for each individual. Bars represent the difference in beliefs across the comparison conditions: INDEPENDENT − CONTESTED *(Top),* INDEPENDENT − DEPENDENT *(Bottom). Dots represent the raw average change in belief towards the direction of each respective consensus condition (independent consensus, dependent consensus, or contested). The plot shows that almost everyone shifted their beliefs more on average for* INDEPENDENT *trials compared to* CONTESTED *trials (top), however, there was substantial variability in how persuaded people were by* INDEPENDENT *trials compared to* DEPENDENT *trials (bottom).*

DEPENDENT trials (negative). Although the majority of participants showed no sensitivity to independence, these results support the idea that there are notable individual differences. For example, at the group level, median belief change was only 2 points higher on INDEPENDENT trials, but for participants best fit by the Alternative model it was 9 points. In addition, the median estimate of belief change for those who were more convinced by a DEPENDENT consensus was just as strong (also 9).

In Supplementary Material 1, we present various exploratory analyses (not preregistered), investigating whether these individual differences are associated with any
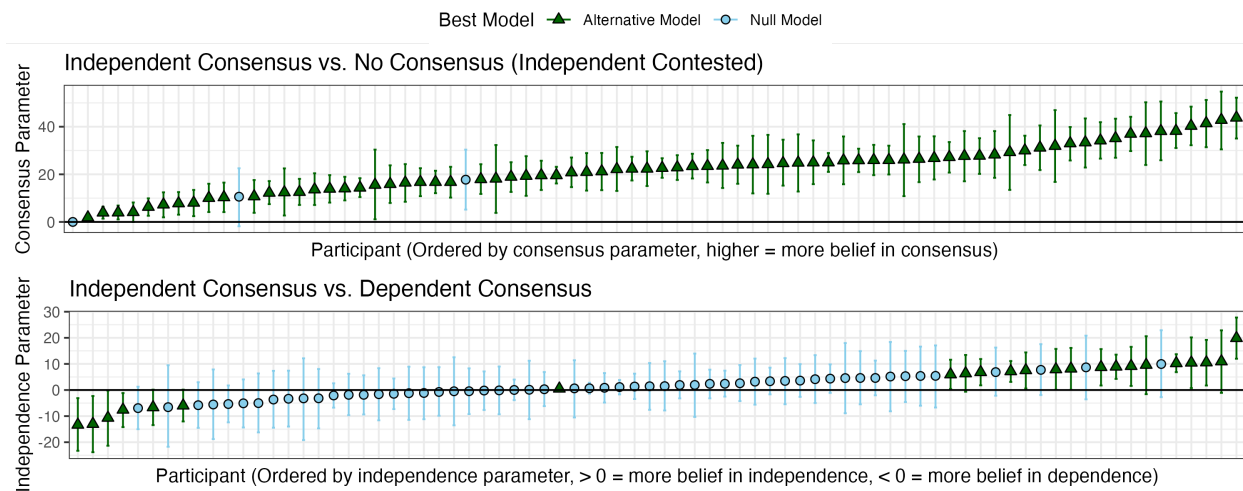

Figure 4

Individual-level modelling results. *(Top panel) For each person (x axis), the estimated persuasiveness of an* INDEPENDENT *consensus relative to* CONTESTED *consensus. The y axis indicates a parameter reflecting the extent to which a participant was more persuaded by a full consensus. The parameter is shown by the point and its 89% credible interval by the lines. Blue circles indicate those participants whose behavior was best captured by a Null model containing only their initial beliefs. Green triangles indicate participants who were more persuaded by consensus. (Bottom panel) For each person, the estimated persuasiveness of an* INDEPENDENT *consensus relative to* DEPENDENT *consensus, with green triangles indicating participants who were sensitive to independence in either direction.*

of the demographic variables we collected (age, education, politics, and social media use). However, our analyses did not reveal any statistical evidence that would allow us to reliably conclude that any of these variables were driving individual differences. However, it is important to emphasize that our study was not designed to test these relationships, so failure to find any credible relationships in our analyses does not mean they do not exist. However, these analyses may be useful for motivating future research interested in more rigorously testing what underlies individual differences in the persuasiveness of different kinds of consensus.

One consideration is that because claims were randomized for each person, different people saw different claims in each consensus condition. Therefore, it is possible that the individual variation we observed could have reflected the fact that some participants were differentially exposed to the claims likely to produce independence effects. One indication of this being an issue would be be if our individual-level results were driven by a small number of trials or claims where participants showed particularly large belief changes. In Supplementary Material 5 we also report a number of follow up analyses (not preregistered) which suggest this individual variation was not due to a few outlier claims and instead reflected consistent preferences for a particular type of consensus.

As an exploratory (not preregistered) analysis of consistency between the two sessions, we also ran the model on each session separately to see if people who were judged as being sensitive to independence overall exhibited similar behavior in both sessions. Although the estimates were substantially noisier because there were fewer trials, there was a strong positive correlation, $r = .57$, $t(21) = 3.07$, $p = .006$, between estimates in the two sessions for participants who were best fit by the Alternative model overall. In other words, people who changed their beliefs more in the direction of INDEPENDENT trials in the first session also did so in the second session, and vice-versa for people who changed more in the direction of DEPENDENT trials. As one might expect,

participants who were classified in neither group did not appear to have any consistent preference across sessions.

Exploratory Analysis: How often did people change their minds?[2]

We have shown that people usually updated their beliefs in line with the consensus, but that the size of this change varied depending on how "knowable" the claim was. What remains unclear is how belief changes varied depending on whether people initially agreed with the claim or not, and how often people actually changed their minds qualitatively, by switching from disbelieving to believing, or vice versa, as opposed to just reducing their certainty in their original belief. Figure 5 shows how much beliefs changed on average based on whether they originally believed the claim (for; belief > 50), disbelieved (against; belief < 50), or neither (equal; belief = 50). Interestingly, there only appeared to be a reasonably clear group-level difference between independent and dependent trials when participants were initially in favor of the claim. In the contested trials, people were more likely to change their beliefs in favor of the claim if they were originally against it, and more likely to revise their beliefs against the claim if they were originally in favor. The directionality of these differences can probably be partially explained by ceiling effects (people had more room to move on the scale in the opposite direction to what they already believed). However, ceiling effects do not explain why the belief revision was more pronounced when participants were originally *for* the claim than when they were initially *against* it.

Another way to explore the same issue, shown in Figure 6, indicates the number of trials where participants qualitatively changed their minds, by either switching from believing to disbelieving, or disbelieving to believing (participants who had a prior rating of 50, indicating no preference, were removed from this analysis). Participants changed their minds on 22% of trials, with 82% of those changes in mind occurring in the consensus conditions. This result shows that not only did some people update their

---

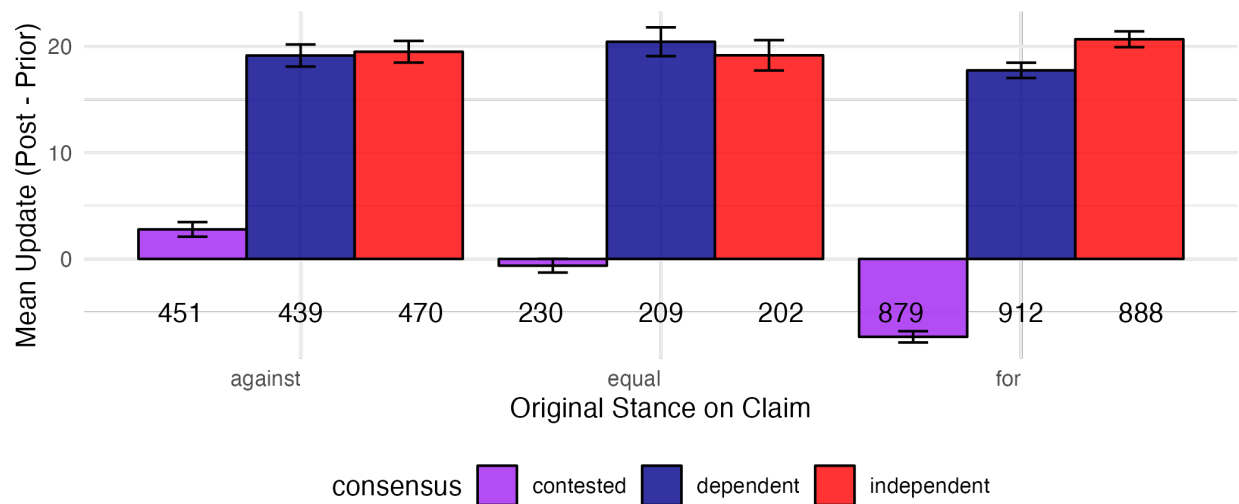[2] None of the analyses that appear in this section were preregistered.

Figure 5

Belief updates as a function of whether the participant originally believed the claim (for), disbelieved (against), or neither (equal). *Numbers represent the number of observations (trials) in each cell. Error bars are standard errors.*

beliefs in line with the consensus, but they actually qualitatively changed their minds. Further, the extent to which people changed their minds corresponded with how knowable the claim claim was: people were more likely to change their minds for knowable claims (7% for knowable eyewitness and 7% for knowable fact) compared to unknowable claims (4% for unknowable fact and 3% for unknowable preference).

Discussion

We performed a large scale assessment of consensus persuasiveness over many different individuals and claim types. Using a mock social media paradigm where each participant assessed the veracity of 60 realistic claims, participants updated their beliefs in light of either 1) an independent consensus (everyone agreeing and citing different sources), 2) a dependent consensus (everyone agreeing but citing the same sources), or 3) no consensus (two in favor, two opposing, all citing different sources). These three trial types allowed us to look at two kinds of consensus effects, which we discuss below.

Full consensus vs no consensus

We were able to replicate previous work demonstrating large group-level belief changes in the direction of claims that are supported by a consensus relative to no consensus (e.g., Ransom et al., 2021). In addition, unlike previous research, we were able to quantify this at the individual level: nearly everybody updated their beliefs in line with the consensus, with substantial variation in how much.

We also found that this effect was stronger for claims that are more "knowable"— that is, claims that are more likely to have a ground truth. While this possibility has been suggested to underlie independence effects (Yousif et al., 2019), we are the first to 1) develop an explicit taxonomy that distinguishes between different kinds of knowability, and 2) show that this taxonomy can reliably predict how persuasive a standard consensus is. This finding makes sense: if it is impossible for anybody to know the truth
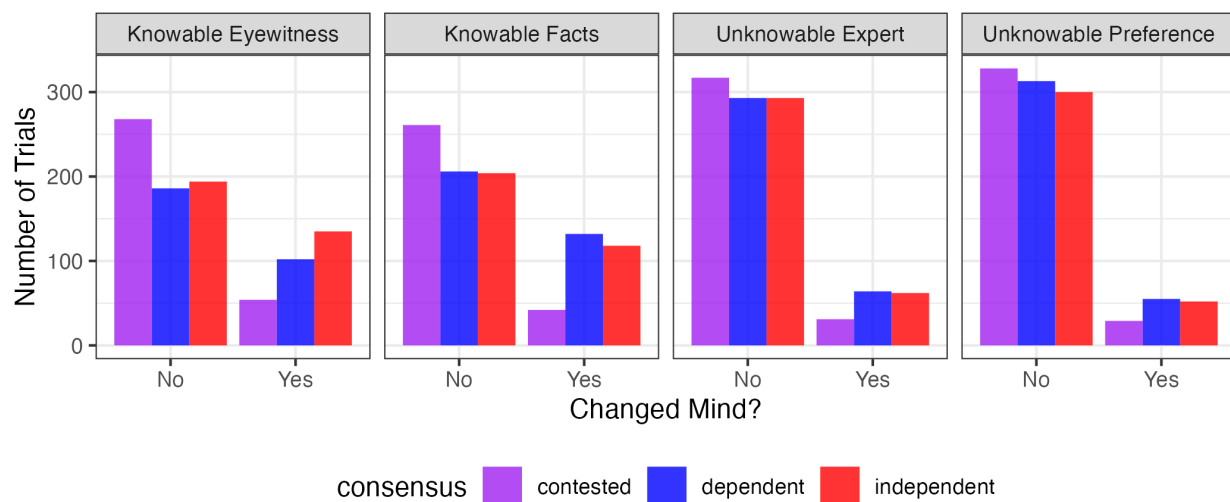

Figure 6

Number of participants who changed their minds (switching from believing to disbelieving, or disbelieving to believing) after viewing the posts. *Trials where participants had an initial rating of 50 (neither believe nor disbelieve) were removed from this analysis (see Figure 5 for the number of trials where participants had no prior preference).*

of a claim, an aggregate of opinions might not be very convincing. Conversely, it might be more sensible to aggregate opinions when there is a knowable ground truth: in such cases, consensus estimates are consistently more reliable than individuals (e.g., Mannes et al., 2014).

The fact that knowability is relevant has important implications. It suggests that if a bad actor wanted to reduce people's belief in a claim that has a consensus among experts, they do not necessarily need to reduce the perception that a consensus exists; all that might be necessary is to create doubt that the claim is knowable at all. Indeed, this kind of rhetoric is already employed by popular climate deniers with wide-reaching platforms such as Jordan Peterson, who have claimed that the climate is too complicated to be modeled accurately (e.g., Readfearn, 2022).

Independent vs dependent consensus

Consistent with a number of recent studies, we found a small effect of consensus independence: at the group level, people tended to be more convinced when everybody cited different sources rather than the same ones (Connor Desai et al., 2022; Simmonds et al., 2023; Yousif et al., 2019). Interestingly, this group-level effect was *not* influenced by the claim type, contrary to the hypothesis posited by Yousif et al. (2019).

At the individual level, most participants in our study were similarly persuaded by an independent consensus compared to a dependent consensus, but the minority that were more persuaded by independence updated their beliefs more than our group-level effects would suggest. Further, a small subset of participants were more convinced by a *dependent* consensus (to an equivalent degree as those who preferred an independent consensus). These individual differences highlight how small or null group-level effects (such as those found in our own study and previous studies, e.g., Alister et al., 2022; Connor Desai et al., 2022; Sulik et al., 2020; Yousif et al., 2019) can be misleading as they hide the heterogeneity in individuals' reliance on a given factor.

The majority who did not prefer an independent consensus is particularly

interesting given that dominant normative, computational theories of reasoning clearly outline that people should be more convinced by independent evidence (Couch, 2022; Madsen et al., 2020; Whalen et al., 2018), or simply assume independence without considering contexts where this may not be the case (e.g., Oktar et al., 2024). Our findings have at least two possible interpretations with respect to these theories. One interpretation of our results, which is how most research has described deviations from normative models in these contexts, is that it shows how people are often irrational when reasoning from a consensus, and do not appropriately weight the informational advantages gained by independence. Our individual-level approach, however, allows for a second interpretation, which is that people had different underlying (but reasonable) assumptions for what independence vs. dependence means. For example, observing multiple people endorsing the same primary source could be a cue to that source's reliability; if that is the case, and source reliability is uncertain enough, a dependence consensus might actually be more persuasive. Variation in the extent to which people make this (or other) assumptions may contribute to the individual variation that we observed. This has important implications for existing theories of consensus reasoning, since it suggests there are reasonable generative assumptions people are making that existing theories do not yet consider.

One consideration is that participants who could not accurately identify the stance of a post were removed. This may have inflated the standard consensus effect relative to the independence effect, if it made people focus more on consensus than they otherwise would have. However, we believe this is unlikely, as our effect size aligns with previous studies that lacked this manipulation check (Alister et al., 2022; Ransom et al., 2021). An open question is whether similarly directing attention to independence would increase sensitivity to it. Additionally, it is also important to caveat that our sample was limited to US participants recruited from Prolific, so our findings may not generalize outside of this sample.

In summary, learning how people are persuaded by different kinds of social consensus is crucial for understanding how information circulating online will affect public opinion (Lewandowsky et al., 2019). The current work used empirical and analytic methods that allowed us to examine the impact of different types of consensus on belief at both the group and individual level. Our finding that individuals vary considerably in their sensitivity to source independence highlights how group-level analyses (which are the norm in this area) can obscure important individual heterogeneity that might be important for illuminating how different people reason about consensus information in the real world.

References

Aboody, R., Yousif, S. R., Sheskin, M., & Keil, F. C. (2022). Says who? Children consider informants' sources when deciding whom to believe. *Journal of Experimental Psychology. General*, *151*(10), 2481–2493. https://doi.org/10.1037/xge0001198

Alister, M., Perfors, A., & Ransom, K. (2022). Source independence affects argument persuasiveness when the relevance is clear. https://doi.org/10.31234/osf.io/5yx36

Anonymous. (Year). Anon title.

Atkisson, C., O'Brien, M. J., & Mesoudi, A. (2012). Adult Learners in a Novel Environment Use Prestige-Biased Social Learning. *Evolutionary Psychology*, *10*(3), 147470491201000309. https://doi.org/10.1177/147470491201000309

Brown, V. A. (2021). An Introduction to Linear Mixed-Effects Modeling in R. *Advances in Methods and Practices in Psychological Science*, *4*(1), 2515245920960351. https://doi.org/10.1177/2515245920960351

Bürkner, P.-C. (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal*, *10*(1), 395. https://doi.org/10.32614/RJ-2018-017

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition [Place: US Publisher: American Psychological Association]. *Journal of Personality and Social Psychology*, *42*(1), 116–131. https://doi.org/10.1037/0022-3514.42.1.116

Center for Countering Digital Hate. (2021). *The Disinformation Dozen* (tech. rep.). Center for Countering Digital Hate.

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

Couch, N. (2022). The diversity principle and the evaluation of evidence. *Psychonomic Bulletin & Review*, *29*(4), 1270–1294. https://doi.org/10.3758/s13423-022-02065-5

Fazio, L. K., Pillai, R. M., & Patel, D. (2022). The effects of repetition on belief in

   naturalistic settings. *Journal of Experimental Psychology: General*, *151*(10),

   2604–2613. https://doi.org/10.1037/xge0001211

Harkins, S. G., & Petty, R. E. (1987). Information utility and the multiple source effect.

   Journal of Personality and Social Psychology, *52*(2), 260–268.

   https://doi.org/10.1037/0022-3514.52.2.260

Harvey, J. A., van den Berg, D., Ellers, J., Kampen, R., Crowther, T. W., Roessingh, P.,

   Verheggen, B., Nuijten, R. J. M., 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

Kruglanski, A. W., Webster, D. M., & Klem, A. (1993). Motivated resistance and

   openness to persuasion in the presence or absence of prior information. *Journal

   of Personality and Social Psychology*, *65*(5), 861–876.

   https://doi.org/10.1037//0022-3514.65.5.861

Kruschke, J. (2014). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*.

   Academic Press.

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

Lin, X., Spence, P. R., & Lachlan, K. A. (2016). Social media and credibility indicators:

   The effect of influence cues. *Computers in Human Behavior*, *63*, 264–271.

   https://doi.org/10.1016/j.chb.2016.05.002

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.
https://doi.org/10.1037/0022-3514.39.2.235

Madsen, J. K., Hahn, U., & Pilditch, T. D. (2020). The impact of partial source
dependence on belief and reliability revision. *Journal of Experimental Psychology:
Learning, Memory, and Cognition*, *46*(9), 1795–1805.
https://doi.org/10.1037/xlm0000846

Mannes, A. E., Soll, J. B., & Larrick, R. P. (2014). The wisdom of select crowds. *Journal
of Personality and Social Psychology*, *107*(2), 276–299.
https://doi.org/10.1037/a0036677

McElreath, R. (2016, January). *Statistical Rethinking: A Bayesian Course with Examples
in R and Stan* [Google-Books-ID: mDo0CwAAQBAJ]. CRC Press.

Mercier, H., & Morin, O. (2019). Majority rules: How good are we at aggregating
convergent opinions? *Evolutionary Human Sciences*, *1*, e6.
https://doi.org/10.1017/ehs.2019.6

Molleman, L., Kurvers, R. H. J. M., & van den Bos, W. (2019). Unleashing the BEAST: A
brief measure of human social information use. *Evolution and Human Behavior*,
40(5), 492–499. https://doi.org/10.1016/j.evolhumbehav.2019.06.005

Oktar, K., & Lombrozo, T. (2025). How aggregated opinions shape beliefs [Publisher:
Nature Publishing Group]. *Nature Reviews Psychology*, 1–15.
https://doi.org/10.1038/s44159-024-00398-7

Oktar, K., Lombrozo, T., & Griffiths, T. L. (2024). Learning From Aggregated Opinion
[Publisher: SAGE Publications Inc]. *Psychological Science*, 09567976241251741.
https://doi.org/10.1177/09567976241251741

Orticio, E., Martí, L., & Kidd, C. (2022). Social Prevalence Is Rationally Integrated in
Belief Updating. *Open Mind*, *6*, 77–87. https://doi.org/10.1162/opmi_a_00056

Pilditch, T. D., Hahn, U., Fenton, N., & Lagnado, D. (2020). Dependencies in evidential

    reports: The case for informational advantages. *Cognition*, *204*, 104343.

    https://doi.org/10.1016/j.cognition.2020.104343

Pillai, R. M., & Fazio, L. K. (2024). Repeated by many versus repeated by one:

    Examining the role of social consensus in the relationship between repetition and

    belief. *Journal of Applied Research in Memory and Cognition*.

    https://doi.org/10.1037/mac0000166

Pucci, V., Kashima, Y., & Perfors, A. (2023). Online communication to the ingroup and

    the outgroup: The role of identity in the "what" and "why" of information sharing.

    Proceedings of the Annual Meeting of the Cognitive Science Society, *45*(45).

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).

Readfearn, G. (2022). 'Word salad of nonsense': Scientists denounce Jordan Peterson's

    comments on climate models. *The Guardian*. Retrieved February 16, 2024, from

    https://www.theguardian.com/environment/2022/jan/27/word-salad-of-nonsense-

    scientists-denounce-jordan-petersons-comments-on-climate-models

Richardson, E., & Keil, F. C. (2022). The potential for effective reasoning guides

    children's preference for small group discussion over crowdsourcing. *Scientific

    Reports*, *12*(1), 1193. https://doi.org/10.1038/s41598-021-04680-z

Sah, S., Moore, D. A., & MacCoun, R. J. (2013). Cheap talk and credibility: The

    consequences of confidence and accuracy on advisor credibility and

    persuasiveness. *Organizational Behavior and Human Decision Processes*,

    121(2), 246–255. https://doi.org/10.1016/j.obhdp.2013.02.001

Simmonds, B. P., Ransom, K. J., & Stephens, R. (2024). Navigating Health Claims on

    Social Media: Reasoning from Consensus Quantity and Expertise. *Proceedings*

of the Annual Meeting of the Cognitive Science Society, *46*(0). Retrieved October 28, 2024, from https://escholarship.org/uc/item/76j8m25k

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

Sulik, J., Bahrami, B., & Deroy, O. (2020). Social influence and informational independence. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *42*.

Toelch, U., Bruce, M. J., Newson, L., Richerson, P. J., & Reader, S. M. (2014). Individual consistency and flexibility in human social information use. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1776), 20132864. https://doi.org/10.1098/rspb.2013.2864

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical bayesian model evaluation using leave-one-out cross-validation and waic. *Statistics and Computing*, *27*(5), 1413–1432. https://doi.org/10.1007/s11222-016-9696-4

Whalen, A., Griffiths, T. L., & Buchsbaum, D. (2018). Sensitivity to Shared Information in Social Learning. *Cognitive Science*, *42*(1), 168–187. https://doi.org/10.1111/cogs.12485

Xie, B., & 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/10.1177/0956797619856844

Zemla, J. C., 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