Limits of repetition in the illusion of consensus

Saoirse Connor Desai [1] (saoirse.connordesai@uts.edu)
Keith Ransom[2] (keith.ransom@unimelb.edu.au)
Laura Angles[3](lauraangles0@gmail.com)
Brett Hayes[3] (b.hayes@unsw.edu.au)
[1]Faculty of Health, University of Technology Sydney
[2]School of Computer & Mathematical Sciences, University of Adelaide
[3]School of Psychology, University of New South Wales

Abstract

Humans rely on social consensus to assess the credibility of information. Confidence in a claim can be influenced by repeated exposure to the same source (dependent consensus), which can create an illusion of consensus. This study investigated whether people differentiate between dependent consensus and consensus derived from multiple independent sources. In two experiments, participants rated their confidence in a claim after exposure to an increasing number of supporting claims within a mock social media environment. Results revealed that although dependent and independent consensus were weighted similarly when exposed to a small number of claims, independent consensus carried greater influence as claim exposures increased, regardless of the claim's valence. These findings were further supported by analysis of free-text justifications of responses using large language models (LLMs). Our findings show that people differentiate between the epistemic weight of different consensus types on social media, where repeated exposure to claims is common.

Keywords: consensus; repetition; source independence; social reasoning; text analysis

Introduction

We rely heavily on others to interpret and explain the world around us. Unfortunately, not everyone we depend on is honest, knowledgeable, or reliable. As misinformation becomes increasingly prevalent, it is more important than ever to understand how we evaluate the credibility of information. One cue that is often used to assess the reliability of a claim is to consider the consensus among supporting opinions (Mercier & Morin, 2019). People are often influenced by widely endorsed ideas, which highlights the powerful impact of consensus on our beliefs and judgments (Lewandowsky, Cook, Fay, & Gignac, 2019; Franzen & Mader, 2023; Ransom, Perfors, & Stephens, 2021).

Although a consensus among various information sources can often suggest that a claim is reliable (Mercier & Morin, 2019), opinions or claims originating from a single source and shared repeatedly by others can create an illusion of widespread agreement. This phenomenon is known as the *illusion of consensus* (Yousif, Aboody, & Keil, 2019). This illusion is illustrated by climate denial blogs. A study by Harvey et al. (2018) found that 80% of 45 science-denying blogs cited a single source, *Polar Bear Science*, which is run by a self-proclaimed expert lacking original research or peer-reviewed publications. Such instances are common (Center for Countering Digital Hate, 2021), highlighting the persuasive impact of repetition.

A social consensus drawn from multiple *independent* information sources is generally seen as more reliable than one based on *dependent* sources, which often repeat others' claims. This principle is enshrined in normative models of judgement and decision-making (Bovens, Hartmann, et al., 2003; Oktar, Lombrozo, & Griffiths, 2024; Whalen, Griffiths, & Buchsbaum, 2018), particularly Bayesian models that prioritize information from independent informants over those relying on shared information (Whalen et al., 2018).

However, empirical studies of how people weigh independent versus dependent consensus have yielded mixed results. Some research indicates that individuals do not differentiate between the epistemic weight of these types of consensus (Aboody, Yousif, Sheskin, & Keil, 2022; Yousif et al., 2019; Sulik, Bahrami, & Deroy, 2020). For example, in an experiment by Yousif et al. (2019), participants read news articles about a claim (e.g., that a new tax policy would boost economic growth). In the independent consensus condition, each article cited a different expert, while in the dependent consensus condition, all articles referenced the same expert. Participants' confidence in the claim increased with both types of consensus compared to a baseline with no consensus. Other studies suggest that people may favor independent consensus slightly more, though the effect is often small. For instance, Alister et al. (2025) presented mock social media posts about various claims (e.g., "narcissists are more politically engaged"). Some posts reflected independent consensus (different experts), while others illustrated dependent consensus (the same expert). A small group-level effect reflected greater confidence in independent consensus. This small effect, however, was carried by a minority of participants - only 23% consistently preferred independent consensus. Connor Desai, Xie, and Hayes (2022) also found that participants were more persuaded by independent than dependent consensus, but only when source independence was strongly emphasized.

Identifying a clear advantage of independent consensus over dependent consensus is difficult, which aligns with the tendency for people to see repeated claims as more credible. An abundant literature has shown that repeated statements are often perceived as more truthful than new, unmentioned ones (e.g., Henderson, Westwood, & Simons, 2022; Fazio, Pillai, & Patel, 2022). There are several possible mechanisms that could underlie this effect. Repetition aids in encoding and

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

familiarity, enhancing retrieval and processing, making repeated claims appear more credible (Alter & Oppenheimer, 2009; Unkelbach, Fiedler, & Freytag, 2007). As noted earlier, people may also infer that information is being intentionally repeated because it comes from a more credible or reliable source (e.g., Colantonio, Durkin, Caglar, Shafto, & Bonawitz, 2021; Oktar & Lombrozo, 2025). Evidence for this mechanism comes from studies that have analyzed participants' written responses to claims based in independent or dependent consensus. Participants exposed to a dependent consensus often perceived a single, frequently cited source as especially credible based solely on its repeated mention (Connor Desai et al., 2022; Connor Desai, Fai, Lee, & Hayes, 2024).

The Limits of Repetition

Current evidence suggests that an overemphasis on repeated information makes it difficult for people to evaluate the strength of claims based on independent versus dependent consensus. However, a significant limitation of many previous studies of consensus effects is that participants were typically exposed to only three or four informants supporting a claim (e.g., Yousif et al., 2019; Connor Desai et al., 2022). In social media contexts, however, individuals often encounter repeated claims much more frequently. This design feature may have amplified the influence of dependent consensus in earlier research.

Recent work shows that repetition can increase the perceived truth of a claim, but this effect diminishes after a certain point. For example, Fazio et al. (2022) found that repeated exposure to true or false trivia statements of up to 16 times over two weeks significantly boosted perceived truthfulness, regardless of the statement's actual truth. However, this effect diminished after 5 to 10 repetitions, with minimal impact from further exposure. Similarly, Hassan and Barber (2021) noted diminishing returns on truth ratings after nine repetitions. In a related paradigm, Ernst, Kühne, and Wirth (2017) identified a "boomerang effect," where excessive repetition of negative messages reduced their perceived credibility. Based on similar findings, Koch and Zerback (2013) suggested that excessive repetition can trigger distrust, making it seem like an attempt to persuade rather than inform.

The Current Studies

Previous research on the illusion of consensus has typically involved limited repetitions from a single source (e.g., four repetitions in Yousif et al., 2019). The current studies reexamine susceptibility to this illusion by exposing participants to either independent or dependent consensus through multiple iterations (up to ten) of fictional social media posts. These posts featured independent expert endorsements, based on the premise that independent consensus would be more persuasive than dependent consensus as exposures increase. We used mock social media posts from fictional news accounts featuring either independent or dependent expert endorsements of a claim (cf. Alister, Ransom, & Perfors, 2022; Ransom et al., 2021). Given the strong normative arguments favoring independent consensus and previous results showing diminishing effects of multiple repetitions (e.g., Fazio et al., 2022), we predicted that people would be more persuaded by independent as compared with dependent consensus as the number of exposures increases.

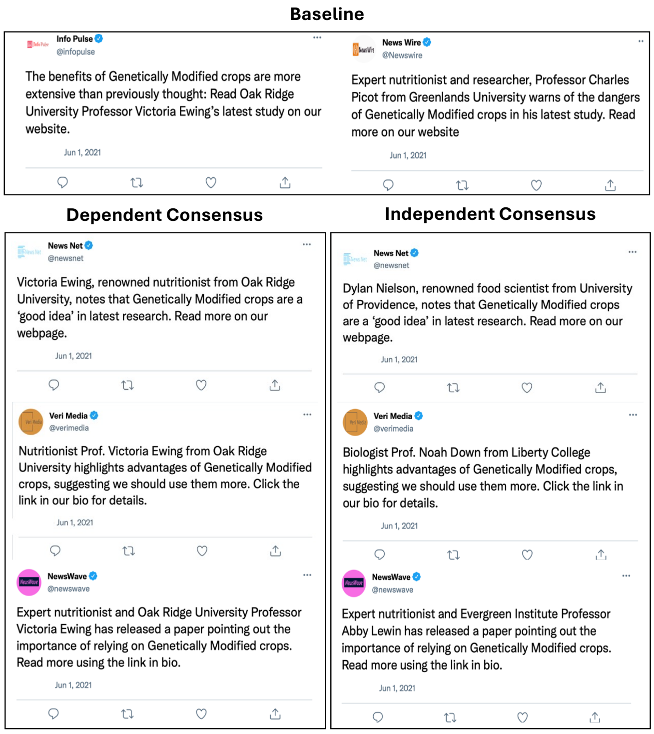

Figure 1: Example Posts in Consensus Conditions Experiment 1.

We also used large language models (LLMs) for exploratory text analyses to categorize participants' justifications for their confidence ratings rather than relying on human coders. Our focus was on how individuals interpreted independent and dependent consensus; specifically whether they differentiated between agreement from independent sources and repetition from a single source.

Experiment 1

Experiment 1 explored the effects of independent versus dependent consensus by presenting a novel claim through multiple social media posts. We aimed to see if up to 10 exposures would increase confidence in claims supported by independent evidence compared to dependent testimony.

Method

Participants In total, 151 participants were recruited via the Prolific crowdsourcing platform and randomly either allocated to an INDEPENDENT or a DEPENDENT CONSENSUS condition. Data from 19 participants (10 in the independent condition and 9 in the dependent condition) were excluded based on post-test understanding of the experimental procedure (final sample: INDEPENDENT CONSENSUS, n = 65; DEPENDENT CONSENSUS = 67). Data from 132 participants were analyzed (M_{age} = 36.88, SD_{age} 12.76, females = 69, males = 59, non-binary = 4). All participants were residents of the US, the UK, Canada, Australia, New Zealand, or Ireland and received £1.50 compensation at the end of the study. The study was not preregsitered.

Table 1: Distribution of Positive (+) and Negative (-) Social Media Posts by Consensus Condition and Phase

		Within subjects							
		Baseline		Phase 1		Phase 2		Phase 3	
		+	-	+	-	+	-	+	-
Between	Dependent	1	1	+3	0	+3	0	+3	0
	Independent	1	1	+3	0	+3	0	+3	0
Cumulative Total		1	1	4	1	7	1	10	1

Materials Experimental stimuli were mock media posts resembling those from the social media platform 'X' (formally known as Twitter) – see Figure 1. Participants were told to imagine that a debate was being held online on the use of genetically modified (GM) crops and that they would see social media posts from news outlets regarding this debate (see OSF for verbatim instructions). We chose this topic because previous studies show that most adults do not have strong prior beliefs about the value of genetically modified (GM) crops (Ransom et al., 2021; Alister et al., 2022). This helps avoid ceiling effects in confidence ratings or floor effects in Experiment 2. We used Anglo-Saxon expert names in the mock social media posts and their academic institutions. Participants in the independent consensus condition viewed posts featuring different experts and institutions.

Participants in the DEPENDENT CONSENSUS condition viewed posts about two experts and academic institutions in the initial phase. One source was repeated in later phases, with posts paraphrased but consistently supporting the claim. Research indicates that paraphrased repetitions can have effects similar to exact repetitions (Pillai & Fazio, 2024). Figure 1 shows examples of each type of consensus stimulus. All posts had identical timestamps and contained no additional sharing or endorsement details, like "likes" or emojis.

All experimental stimuli were produced using TWEETGEN Beta. Media site names were generated using Chat GPT-3.5 (2024), and media site logos were generated using LOGO.com. See OSF for full set of stimuli.

Procedure The experiment was coded and run using JSPsych (De Leeuw, 2015). Participants provided consent and demographic information before reading instructions, followed by a three-item comprehension test. A perfect score was required to proceed; failure meant starting over.

Participants viewed social media posts across four phases, as summarized in Table 1. We presented all participants with two posts sequentially at baseline; one supporting and one opposing the use of GM crops, with order randomized. They had unlimited time to read each post, but could advance to the next screen 3 s after a post appeared. After both posts, they rated their confidence with the claims supporting GM crops, "Based on all the claims you have seen so far, to what extent do you agree that genetically modified crops are a good idea?", using a 0-100 on screen slider (0 = Strongly disagree, 50 = Neutral, 100 = Strongly agree).

In the three subsequent phases of the study, participants were shown three sequential social media posts, all supporting the use of genetically modified (GM) crops. The presentation method mirrored that of the baseline phase. After each set of posts, participants re-rated their confidence in the claim. Those in the INDEPENDENT CONSENSUS condition viewed 10 social media posts over four phases reporting that different experts supported the claim. A different combination of expert name and academic institution was used for each post, with a different media outlet responsible for each post. Those in the DEPENDENT CONSENSUS condition viewed the same number of posts, all reporting that the same expert supported the claim. The name of the expert and academic institution remained the same for each post, but a different media outlet was responsible for the post. The particular expert/institution names used in the dependent condition, were randomly selected for each participant from the nine possible combinations.

After phase 4, participants completed two memory checks for the post details. They estimated the total number of positive claims and total number of positive negative claims presented during the main task. They were then presented with a randomly ordered list of 14 expert names, containing the 11 names or 2 names used in the INDEPENDENT or DEPENDENT conditions respectively. They were asked to click on all those that actually were presented during the task.

Participants then completed several post-test questions. They rated their own view of GM foods prior to participation. They were next asked to type a free text response to a question about the effects of consensus on their belief in the claim. This was worded differently for the INDEPENDENT and DEPENDENT conditions "Most of the posts you saw cited [different] / [the same] experts. How did this influence your judgment about whether genetically modified crops are a good idea?". They were then asked to respond to the counterfactual question "How would your judgment have changed (if at all) if all the posts had cited [the same] / [different] experts?" They were also asked to rate their willingness to share one of the posts that had a positive view of GM foods. Responses to this question showed that people were generally unwilling to share (only 30.6 percent responded positively) and are not analyzed further in this paper (see OSF for further details).

Exclusion Criteria Participants in the INDEPENDENT CONSENSUS condition were excluded if they selected only one expert name during the memory check, if all selected names were incorrect, or if their estimates of positive and negative claims were negative or zero. In the DEPENDENT CONSENSUS, participants were excluded for selecting more than two different expert names or names differing from those presented in the experiment, as well as for negative differences between their positive and negative estimates (a zero difference was acceptable, indicating a response to the number of *independent* expert posts viewed).

Results & Discussion

All analyses were conducted using JASP (Version 0.18.3; JASP Team, 2025). Target claim confidence was analyzed with Bayesian mixed analysis of covariance (BANCOVA). Default prior settings were applied unless otherwise specified, and all models included random slopes for repeated measures factors. Model comparisons were based on Bayes factors and posterior probabilities to identify the best-fitting model. Following Lee and Wagenmakers (2014), a Bayes Factor (BF_{10}) of 1 to 3 indicates "anecdotal" evidence for the alternative hypothesis, 3 to 10 represents "moderate," 10 to 30 shows "strong," 30 to 100 indicates "very strong," and above 100 signifies "extreme" evidence. For the null hypothesis, a BF_{10} between 0.33 and 1 indicates anecdotal evidence, 0.33 to 0.1 indicates moderate, and 0.1 to 0.03 represents strong.

Target claim confidence Figure 2 shows that confidence increased from Baseline to Phase 1 in both consensus conditions. In subsequent phases, ratings diverged showing increased confidence in INDEPENDENT CONSENSUS but no change in DEPENDENT CONSENSUS. The BANCOVA model, which included *Prior Beliefs* as a covariate, *Phase*, *Consensus*, and their interactions, yielded the highest posterior probability ($P(M \mid \text{data}) = 0.810$), indicating strong support for its explanatory power ($BF_M = 38.275$).

The analysis of effects provided evidence for the inclusion of several predictors. The effect of *Phase* showed extreme evidence in favor of inclusion, with $P(\text{incl} \mid \text{data}) = 0.190$ ($BF_{\text{incl}} = 1.574 \times 10^{40}$). The interaction between *Phase* and *Consensus* demonstrated moderate evidence for inclusion, with $P(\text{incl} \mid \text{data}) = 0.810$ and $BF_{\text{incl}} = 6.663$. *Prior Beliefs* had extreme evidence for inclusion, as reflected by $P(\text{incl} \mid \text{data}) = 1.000$ and $BF_{\text{incl}} = 2.006 \times 10^9$. In contrast, the effect of Consensus alone provided only anecdotal evidence for inclusion, with $P(\text{incl} \mid \text{data}) = 0.116$ and $BF_{\text{incl}} = 1.671$.

Experiment 1 revealed that confidence evolves differently across multiple exposures to claims based on the type of consensus. In both consensus conditions confidence initially increased. Across subsequent exposures, however, confidence patterns for each condition diverged. In the DEPENDENT condition, confidence stabilized, showing that repeated agreement among the same individuals did not further increase confidence. In the INDEPENDENT condition, on the other hand, confidence continued to grow, indicating that new, independent sources of agreement bolster confidence in the target claim. This suggests that the impact of consensus depends on both agreement and the independence of contributors.

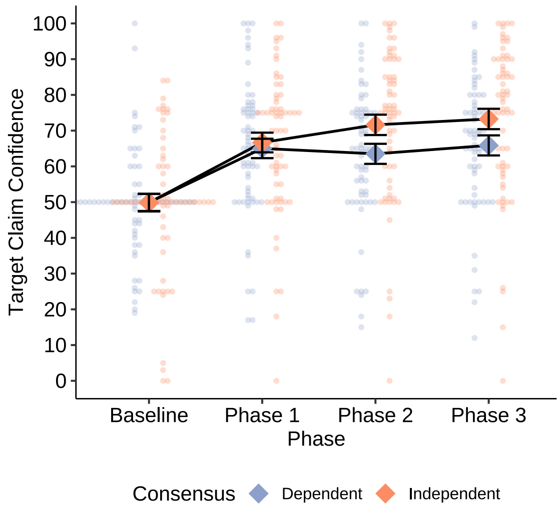

Figure 2: Target claim confidence as a function of consensus type in Experiment 1. Diamonds represent the mean and error bars represent 95% confidence intervals. Jitter represents individual data points.

Experiment 2

Experiment 1 confirmed our hypothesis that independent consensus is more persuasive than dependent consensus when people encounter more of the same claim in increasing numbers. Experiment 2 explored whether a similar pattern would occur when the consensus valence was reversed, specifically focusing on negative claims about GM foods. While studies of illusory truth suggest that the perceived truth of a repeated claim is not influenced by its valence (Unkelbach, Bayer, Alves, Koch, & Stahl, 2011), other research indicates that message valence can affect both the frequency and speed of sharing content on social media (Tsugawa & Ohsaki, 2015).

Method

Participants A total of 162 participants were recruited and randomly allocated to either an INDEPENDENT or a DEPENDENT CONSENSUS condition. The majority (n = 150) were recruited and tested online via Prolific and received the same monetary payment as in the previous experiment. A minority (n = 12) were recruited from the UNSW undergraduate subject pool and received course credit for participation. Preliminary analyses found no differences between the ratings patterns of these subgroups, so they were combined. Using the same criteria as Experiment 1, data from 16 participants was excluded (10 in the independent condition, 6 in the dependent condition) (final sample: INDEPENDENT CONSENSUS, n = 74; DEPENDENT CONSENSUS = 72). Data from 146 participants was analyzed ($M_{age} = 38.16$, $SD_{age} = 14.68$, females = 74, males = 72). The study was preregistered – see `https://osf.io/q379s/?view_only=f5e742f824334386b4326fee1ac4a1a9` for details.

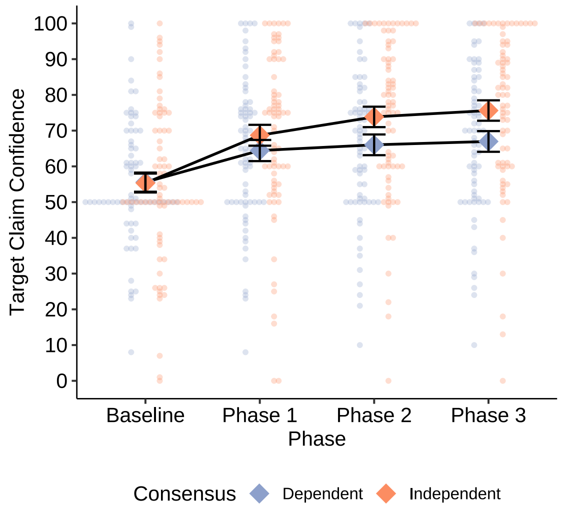

Figure 3: Target claim confidence as a function of consensus type in Experiment 2. Diamonds represent the mean and error bars represent 95% confidence intervals. Jitter represents individual data points

Design, Materials & Procedure The experimental design and procedure were the same as Experiment 1 with the following exceptions. The most important change was in the valence of the claims and ratings concerning GM foods. In this case, most posts in both consensus conditions reported expert views *opposing* the use of GM crops. We accordingly modified the rating required in each phase, "Based on all the claims you have seen so far, to what extent do you agree that genetically modified crops are a bad idea?". The other change was that the question about prior views on GM foods was presented *before* the main task.

Results & Discussion

Figure 3 shows that the pattern of change in subsequent ratings was generally similar to that observed in the previous study. The BANCOVA results revealed that the best-fitting model, which included *Phase*, *Prior Beliefs*, *Consensus*, and the interaction *Phase* × *Consensus*, yielding the highest posterior probability ($P(M \mid \text{data}) = 0.921$), and indicating strong support for its explanatory power ($BF_M = 104.248$).

The analysis of effects revealed varied levels of evidence for inclusion across the predictors. The effect of *Phase* showed overwhelming evidence for inclusion, with $P(\text{incl} \mid \text{data}) = 0.079$ and $BF_{\text{incl}} = 8.983 \times 10^{33}$. The variable *Prior Beliefs* demonstrated extreme evidence for inclusion, as $P(\text{incl} \mid \text{data}) = 1.000$ and $BF_{\text{incl}} = 4.917 \times 10^4$. In contrast, the effect of *Consensus* showed anecdotal evidence for the null hypothesis, with $P(\text{incl} \mid \text{data}) = 0.024$ and $BF_{\text{incl}} = 0.435$. The interaction term, *Phase* × *Consensus*, showed strong support for inclusion, with $P(\text{incl} \mid \text{data}) = 0.921$ and $BF_{\text{incl}} = 38.250$.

As in Experiment 1, the interaction between Phase and Consensus again suggests that differentiation between independent and dependent consensus increased with successive exposures to independent or dependent (repeated) claims.

Exploratory Text Analysis

To further examine participants' preferences regarding independent versus dependent consensus, we used two large language models (LLMs) – Anthropic's "Claude 3.5 Sonnet" (model released October 22, 2024) and Google's Gemini 1.5 Pro (Georgiev et al., 2024) – to analyze their free-text responses to two of the post-test questions. The prompts were thus specific to each participant and included the claim about genetically modified crops, the text of the two questions, and both of the participant's responses. Due to their considerable length, full versions of the prompts and responses are, full details of LLM prompts, the response coding system and LLM instructions to rate the difficulty of each coding task are available in our OSF repository [1].

Responses to the first question regarding how multiple posts citing either different experts (or the same expert, depending on condition) influenced beliefs, were coded: as "positive" or "negative" if the nature of the consensus had a positive or negative impact on their belief, "none" if it did not, "other positive" or "other negative" if other factors were discussed instead, or "unspecified" if no better label applied. Responses to the second question, which asked people whether they would have been more convinced by the consensus they saw or the relevant counterfactual, was coded to capture the preference expressed: "dependent", "independent", "none" (i.e., no preference for either), or "unspecified". For each question the models rated the difficulty of the coding task as: "easy" if the answer was explicitly stated, "moderate" where it was implied, "difficult" where significant guess work was involved, or "impossible" for poor quality responses.

We analyzed the effect of three binary factors on the generated response codings: LLM (CLAUDE vs GEMINI), consensus condition (DEPENDENT vs INDEPENDENT), and experiment (EXP1 vs EXP2). Response proportions for the respective coding options for each question are shown in Figure 4, which reveals two important patterns consistent with our overall findings. Firstly, according to our LLM raters, people in the INDEPENDENT CONSENSUS condition reported that the expert citations they saw had a positive impact on their confidence in the claim presented, more frequently than people in the DEPENDENT condition ($\chi^2(1) = 31.9, p < .001$). Likewise, they reported a negative impact less frequently ($\chi^2(1) = 22.6, p < .001$). Secondly, when asked to directly consider the counterfactual condition, people's responses indicated a strong preference in favor of an independent consensus over a dependent one ($\chi^2(1) = 58.5, p < .001$)[2].

[1] As this analysis was exploratory, and because our checks revealed plausible responses from the LLM's, prompt/parameter sweeps and ablation studies were deferred for future work.

[2] Wald χ^2 tests were based on a multinomial logistic regression model predicting the relevant response coding on the basis of the three binary factors discussed (with no interactions).

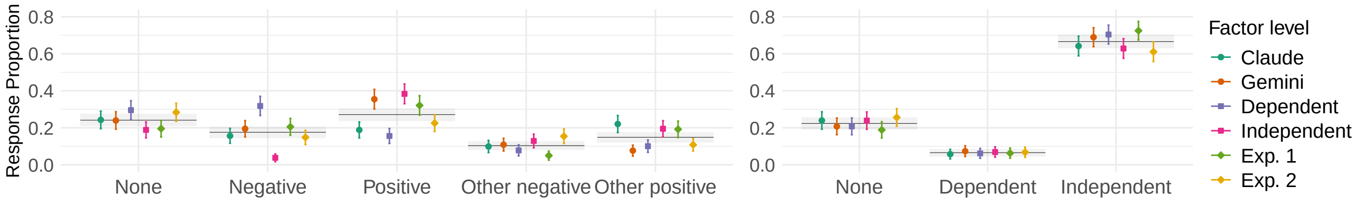

Figure 4: Proportion of response codings generated by two LLMs for people's free-text responses to questions concerning the influence of the expert citations they saw (left panel), and the kind of consensus they prefer (right panel). Solid horizontal lines represent mean proportions per response type, coloured points represent marginal means across three factors (different shapes) at each of two levels (same shapes). Error bars and shaded regions represent 95% CI of the respective mean. The infrequent "unspecified" responses are omitted to save space.

General Discussion

We investigated whether people are sensitive to the epistemic basis of different consensus types with increasing exposures to an expert claim informed by one expert (DEPENDENT CONSENSUS) or multiple different experts (INDEPENDENT CONSENSUS). Both experiments demonstrated that, after a limited number of claim exposures, participants gave equal weight to both consensus types, aligning with prior research indicating insensitivity to how a consensus is formed with a small number of claim exposures (Yousif et al., 2019; Xie & Hayes, 2022). A key novel contribution was our finding that confidence in claims based on independent or dependent consensus diverged with increased exposure to the claims. Confidence in the independent consensus continued to rise, whereas confidence in the dependent consensus leveled off. This trend aligns with the existing illusory truth literature, which posits that extensive repetitions can erode belief in repeated claims (Hassan & Barber, 2021; Fazio et al., 2022). Experiment 2 replicated these results with negatively valenced claims, thereby reinforcing the robustness of our findings.

The differential levels of confidence observed indicate that individuals do indeed respond to the type of consensus as its dependency becomes increasingly apparent through increased exposure. This finding challenges prior assertions that individuals generally exhibit insensitivity toward different consensus types (Yousif et al., 2019). To our knowledge, this study is the first to demonstrate that individuals can effectively distinguish between consensus types following extensive exposure, thus advancing the discourse surrounding the illusion of consensus and sensitivity to independence between sources.

Another novel contribution of this study is exploratory analysis of participant's free-text confidence justifications with two LLMs. Findings revealed two clear patterns. Participants in the independent consensus condition were more likely to report that expert citations positively influenced their confidence in the presented claim compared to those in the dependent consensus condition. They were also less likely to indicate a negative influence. Furthermore, when asked to evaluate the counterfactual condition, participants strongly favored independent consensus over dependent consensus. These patterns were consistent across both experiments and highlight the preference for, and perceived reliability of, independent expert agreement. Although preliminary, these findings provide further evidence that individuals have different interpretations of dependent and independent consensuses, and are consistent with similar categorisation of free-text responses using human coders (Connor Desai et al., 2022, 2024).

It is important to note the constraints on interpretation of these results. Most notably, our findings are based on a single claim where participants do not have strong prior beliefs about its truth or falsity (Ransom et al., 2021). Future research should examine whether these patterns hold for a broader range of claims, particularly those with polarized opinions. Increased exposure may lead to different interpretations for claims with more skewed distributions of prior belief.

Conclusions

Our findings indicate that individuals are sensitive to how consensus is formed and become more aware of its structure with repeated exposure. These results challenge the idea of general insensitivity to consensus formation and support the notion that excessive repetition can weaken belief in repeated claims. The findings provide valuable insights for domains such as public health or science communication where recipients' perceptions of consensus impact trust and engagement. Acknowledging that repeated exposure can weaken confidence in certain consensus types emphasizes the need for thoughtful messaging in contexts where there is likely to be repeated exposure to a claim.

Acknowledgements

This research was funded by an ARC Discovery Grant (DP 220101592) awarded to BK. We thank Won Jae Lee for his assistance with the programming of the experimental task.

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.

Alister, M., Ransom, K. J., Connor Desai, S., Soh, E. V., Hayes, B., & Perfors, A. (2025). How convincing is a crowd? quantifying the persuasiveness of a consensus for different individuals and types of claims. *Psychological Science*. (Accepted for publication)

Alister, M., Ransom, K. J., & Perfors, A. (2022). Source independence affects argument persuasiveness when the relevance is clear. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 44).

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and social psychology review*, *13*(3), 219–235.

Bovens, L., Hartmann, S., et al. (2003). *Bayesian epistemology*. Oxford University Press on Demand.

Center for Countering Digital Hate. (2021). *The Disinformation Dozen* (Tech. Rep.). Center for Countering Digital Hate. Retrieved from `https://www.counterhate.com/disinformationdozen`

Colantonio, J., Durkin, K., Caglar, L. R., Shafto, P., & Bonawitz, E. (2021). The intentional selection assumption. *Frontiers in psychology*, 4664.

Connor Desai, S., Fai, J., Lee, J., & Hayes, B. (2024, October 29). *Explaining away the illusion of consensus.* Retrieved from `https://doi.org/10.31234/osf.io/9hnm7` (PsyArXiv Preprint)

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

De Leeuw, J. R. (2015). jspsych: A javascript library for creating behavioral experiments in a web browser. *Behavior research methods*, *47*, 1–12.

Ernst, N., Kühne, R., & Wirth, W. (2017). Effects of message repetition and negativity on credibility judgments and political attitudes. *International Journal of Communication*, *11*, 21.

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.

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

Georgiev, P., Lei, V. I., Burnell, R., Bai, L., Gulati, A., Tanzer, G., ... Vinyals, O. (2024). *Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context.* Retrieved from `https://arxiv.org/abs/2403.05530`

Harvey, J. A., Van Den Berg, D., Ellers, J., Kampen, R., Crowther, T. W., Roessingh, P., ... others (2018). Internet blogs, polar bears, and climate-change denial by proxy. *BioScience*, *68*(4), 281–287.

Hassan, A., & Barber, S. J. (2021). The effects of repetition frequency on the illusory truth effect. *Cognitive research: principles and implications*, *6*(1), 38.

Henderson, E. L., Westwood, S. J., & Simons, D. J. (2022). A reproducible systematic map of research on the illusory truth effect. *Psychonomic Bulletin & Review*, 1–24.

JASP Team. (2025). *JASP (Version 0.19.3)[Computer software].* Retrieved from `https://jasp-stats.org/`

Koch, T., & Zerback, T. (2013). Helpful or harmful? how frequent repetition affects perceived statement credibility. *Journal of Communication*, *63*(6), 993–1010.

Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge university 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. doi: 10.3758/s13421-019-00948-y

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

Oktar, K., & Lombrozo, T. (2025). How aggregated opinions shape beliefs. *Nature Reviews Psychology*, 1–15.

Oktar, K., Lombrozo, T., & Griffiths, T. L. (2024). Learning from aggregated opinion. *Psychological Science*, 09567976241251741.

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*.

Ransom, K. J., Perfors, A., & Stephens, R. (2021). Social meta-inference and the evidentiary value of consensus. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 43).

Sulik, J., Bahrami, B., & Deroy, O. (2020). Social influence and informational independence. In *Cognitive science conference proceedings* (Vol. 19).

Tsugawa, S., & Ohsaki, H. (2015). Negative messages spread rapidly and widely on social media. In *Proceedings of the 2015 acm on conference on online social networks* (pp. 151–160).

Unkelbach, C., Bayer, M., Alves, H., Koch, A., & Stahl, C. (2011). Fluency and positivity as possible causes of the truth effect. *Consciousness and cognition*, *20*(3), 594–602.

Unkelbach, C., Fiedler, K., & Freytag, P. (2007). Information repetition in evaluative judgments: Easy to monitor, hard to control. *Organizational Behavior and Human Decision Processes*, *103*(1), 37–52.

Whalen, A., Griffiths, T. L., & Buchsbaum, D. (2018). Sensitivity to shared information in social learning. *Cognitive science*, *42*(1), 168–187.

Xie, B., & Hayes, B. (2022). Sensitivity to evidential dependencies in judgments under uncertainty. *Cognitive Science*, *46*(5), e13144.

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.