

### Abstract

Consensus between informants is a valuable cue to a claim's epistemic value, when informants' beliefs are developed independently of each other. Recent work (Yousif et al., 2019) described an *illusion of consensus* such that people did not generally discriminate between the epistemic warrant of *true consensus*, where a majority claim is supported by multiple independent sources, and *false consensus* arising from a single source's repeated claim. Four experiments tested a novel account of the *illusion of consensus*; that it arises when people are unsure about the independence of the primary sources on which informant claims are based. When this independence relationship was ambiguous we found evidence for the illusion. However, when steps were taken to highlight the independence between data sources in the *true consensus* conditions, and confidence in a claim was measured against a *no consensus* baseline (where there was an equal number of reports supporting and opposing a claim), we eliminated the *illusion of consensus*. Under these conditions, more weight was given to claims based on *true consensus* than *false consensus*. These findings show that although the *illusion of consensus* is prevalent, people do have the capacity to distinguish between *true* and *false consensus*.

*Keywords:* Judgment, Reasoning, Dependency, Consensus

### Getting to the Source of the Illusion of Consensus

Many everyday judgments and decisions are made by evaluating and integrating reports from multiple informants. One cue used to guide our confidence in such reports is the degree of consensus between them. Consensus can be a valuable cue to a claim's epistemic value and generally increases the amount of evidence available. However, the epistemic value of consensus is typically reduced when reports are informationally dependent (Dietrich & Spiekermann, 2013; Whalen et al., 2018). A set of reports can be considered independent if each informant considers data and makes conclusions in isolation from all other informants. Reports are rendered dependent when informants share data or conclusions (Hahn et al., 2019; Schum, 2001), or background (Bovens, Hartmann, et al., 2003; Madsen et al., 2020). In an extreme case of dependence the same information may be repeated by the same individual (Weaver et al., 2007; Xie et al., 2020).

When booking accommodation for an upcoming trip, for example, we might consult a review website and read reviews from a variety of previous guests. If different guests independently provide similarly positive reviews, this constitutes a *true consensus*. We might, therefore, choose the option that has the most positive reviews. Conversely, we might notice that several reviews were written by the same person, or members of the same family, causing us to give less weight to those reviews. This latter case represents a *false consensus*<sup>1</sup>.

Failing to detect a *false consensus* can have serious implications for accurate belief formation; for instance, over 80% of climate change denial blogs repeat claims from a single informant (Harvey et al., 2018). The ability to form accurate beliefs based on converging reports also has crucial implications for adhering to public health advice during a global health crisis. A single informant's false or misleading claim can be reiterated via news and

---

<sup>1</sup> The *illusion of consensus* that we refer to here should to be differentiated from the *false consensus effect* (Ross et al., 1977) which describes the egotistical bias that one's personal qualities, beliefs, and actions are widespread in the population.

social media, increasing its fluency and perceived truthfulness (DiFonzo et al., 2016; Fazio et al., 2015; Schwarz et al., 2007; Weaver et al., 2007). The ability to discriminate between *true* (independent) and *false* (dependent) consensus therefore appears to be critical to the accurate evaluation of informant claims.

The current literature on whether people are sensitive to informational dependencies when making consensus-based judgments offers conflicting accounts. Several studies suggest that adults are able to discount *false consensus* when dependent reports are explicitly presented as literal repetitions of previous reports (e.g., Xie et al., 2020). Children are similarly able to recognize the redundancy of repeated reports (e.g., Kim & Spelke, 2020), and by age six endorse claims supported by a greater number of primary sources (Aboody et al., 2021).

There is also evidence that adults can exploit simple cues to dependency. Mercier and Miton (2019) examined sensitivity to different types of information dependency in consensus-based judgments by presenting participants with short vignettes describing informant's opinions; varying the number of informants (1 vs. 3) and the degree of dependency between informant's opinions. Participants in one experiment were asked to rate their preference for a restaurant after reading a vignette in which informants' opinions about the restaurant were either independent of one another or relied on a common third party through hearsay. Participants gave less weight to the opinions of three informants when the opinion of three informants relied on the same third party. In further experiments, they found that participants discounted the value of consensual reports when informants' opinions had a common motivation, or when the informants were perceived as untrustworthy.

The view that people readily discount reports based on false or informationally dependent consensus, however, has been challenged. In a series of studies by Yousif, Aboody and Keil (2019, henceforth YAK), participants read four texts (fictitious student essays or news articles) that took a positive position (e.g., that Sweden's tax policy should

be supported), and one that took a negative position (e.g., Sweden's tax policy should not be supported). Each article cited a unique primary source (e.g., a different economic foundation). In the *true consensus* condition four different sources were cited as supporting the claim and one source opposed the claim. The *false consensus* condition followed the same structure except that all of the supporting claims came from a single source (i.e., a single economic foundation). There was also a *no consensus* baseline, where there was one supporting and one opposing text. Crucially, the number of sources cited in the *no consensus* and *false consensus* conditions were equivalent.

YAK found an *illusion of consensus* such that participants gave similar confidence ratings for claims made by different primary sources (*true consensus*) and the repeated claim of a single primary source (*false consensus*). This failure to give more weight to *true consensus* could not be explained by a failure to notice the difference between sources, as excluding people with poor source memory did not eradicate the effect. The *illusion of consensus* result was robust across four experiments (with one exception that we address in the General Discussion). Notably, no difference was found between confidence in claims based on *false* and *true consensus* even when participants explicitly stated that, in principle, *true consensus* should carry more weight than *false consensus*.

YAK also found that confidence ratings were higher in both the *false* and *true consensus* conditions compared to the *no consensus* condition. This result is consistent with many previous findings showing that mere repetition of a claim often increases its credibility (e.g., Dechêne et al., 2010; Henkel & Mattson, 2011; Kim & Spelke, 2020; Koch & Zerback, 2013). Such effects of information repetition are likely to involve processes such as increased familiarity or accessibility of the repeated information (Begg et al., 1992; Weaver et al., 2007). Some participants may also believe that repetition of information implies that the informant has intentionally selected that information because it is especially credible or useful (Colantonio et al., 2021; Estlund, 1994). These mechanisms may also contribute to the *illusion of consensus* by boosting confidence in claims in a *false*

*consensus* condition.

### Source dependence and the illusion of consensus

YAK's findings of susceptibility to an *illusion of consensus* are compelling because they found no difference in confidence ratings between *true* and *false consensus* conditions across multiple experiments with different sets of stimulus materials. These results are also important because the YAK procedure contains features that map onto contexts outside the laboratory where people rely on information from informants to evaluate claims (e.g., public health recommendations, environmental advice). As is common in such contexts, the claims that participants had to judge in the YAK studies involved evaluating opposing majority and minority views by authorities who are assumed to have more expert knowledge than the participants (cf. Rabb et al., 2019). Hence, the YAK *illusion of consensus* results have already had considerable impact - being cited as an explanation for the persistence of beliefs in fake news and anti-science views (e.g., Kozyreva et al., 2020; Schwarz & Jalbert, 2020).

However, given the conflicting results reported in studies like YAK and Mercier and Miton (2019), it is crucial to gain a better understanding of the cognitive mechanisms that give rise to the *illusion of consensus*. We argue that sensitivity to the relative epistemic value of *true* and *false consensus* depends on one's understanding of the causal dependence or independence between the primary data sources on which informant claims are based. As alluded to earlier, one potential explanation for the *illusion of consensus* is that people are not convinced that reports in the *true consensus* condition are truly independent. For example, in the YAK study where *true consensus* reports come from ostensibly separate economic advisory bodies, people may assume that the bodies relied on a common data source and/or methodology to arrive at their conclusions. This seems like a reasonable assumption given the view that dependence between information sources is more common in the real-world than independence (Kruskal, 1988; Navon, 1978; Winkler & Murphy,

1973).

Our chief aim, therefore, was to test an account of the *illusion of consensus* that accords a central role to people’s understanding of the independence of the informants’ data sources. When they are unsure about the independence of this data, the illusion is likely to arise, with similar weight being given to claims based on *true* and *false consensus*. However, when these dependence relationships are made transparent, the illusion should be eliminated.

The causal and statistical dependence between information sources is a key factor in formal models of belief revision based on social evidence (Anderson & Holt, 1997; Whalen et al., 2018). Such models predict that less weight should be attached to the information provided by multiple informants when that information is causally dependent (i.e., informants obtain their evidence from a common source). This prediction has been supported in studies of intuitive probabilistic judgments (Whalen et al., 2018), and propensity to follow the advice given by multiple informants (Yaniv et al., 2009). However, ours is the first study to examine whether the perceived causal dependence between information sources is the basis for the *illusion of consensus*.

As noted earlier, the understanding of causal relations between sources in the *true consensus* condition may not be the only factor contributing to the *illusion of consensus*. Increased confidence in claims in a *false consensus* condition may occur through mere repetition, or a perception that repeated sources are more credible. We suggest, however, that an understanding of the dependency relations between sources is central to the *illusion of consensus*, such that if these relations are made more transparent, the illusion will be reduced or eliminated.

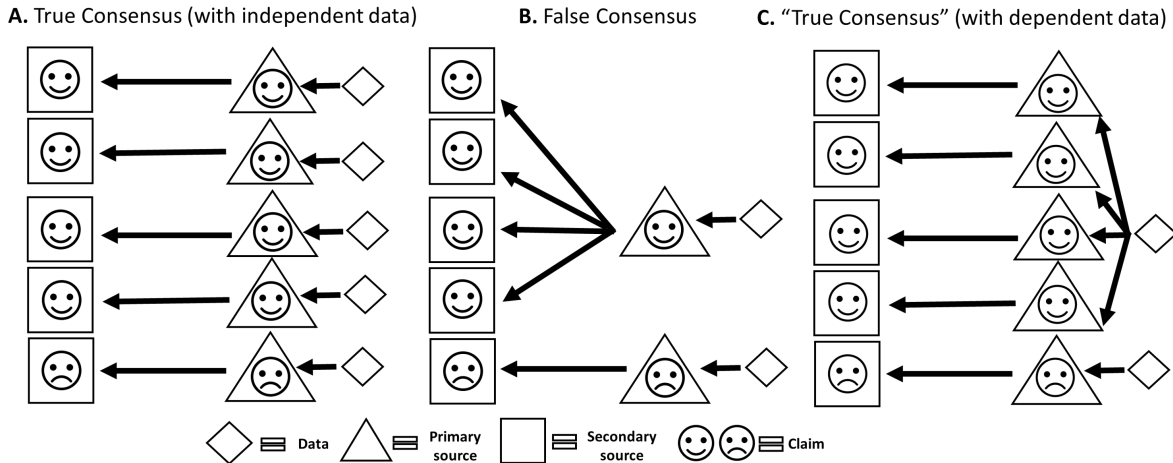
## Overview of Experiments

The goal of our experimental program was to examine how the transparency of the dependency relations between the information provided in each report affects the *illusion*

*of consensus*. We first clarify our theoretical framework and the terminology used throughout the rest of the paper. In this framework, “data” refers to the information that a primary source uses to formulate their claim. The secondary source refers to whoever reports the primary source’s claims. Putting this into the context of the Swedish tax scenarios examined in YAK, the data (e.g., survey results, economic statistics) are used by primary sources (economic foundations) to derive a claim (e.g., the tax policy is a net positive), and the secondary source is the student essay citing the economic foundation’s arguments. Figure 1 illustrates the core difference in the relations between primary and secondary sources in the *true* and *false consensus* conditions.

In studies which have found evidence for the *illusion of consensus* such as YAK, participants may believe that the primary sources in the *true consensus* condition are actually dependent because they rely, to some degree, on a common data set and/or similar methodologies (as illustrated in Figure 1C). For example, when economists have access to the same national statistics or are educated via the same school of thought (cf. Madsen et al., 2020). Such a causal dependency amongst reports would mean that the ostensibly “independent” reports in the *true consensus* condition were perceived as different versions of the same primary sources of evidence. If this was the case, then this would explain the failure to give these reports more epistemic weight than those in the *false consensus* condition.

We hypothesize that the *illusion of consensus* arises when judgments based on *false consensus* (see Figure 1B) are compared with those based on *true consensus* with primary source data that is perceived as dependent (see Figure 1C). The general paradigm that we used to test this claim was similar to that used by YAK, in that participants were asked to rate their confidence in a claim after receiving informant information under conditions that corresponded to YAK’s *true*, *false* and *no consensus* conditions. Crucially, however, we examined the role of perceived dependence between information sources by either adding an explicit statement about the causal dependence or independence between primary



**Figure 1**

*Schematic diagram representing informational dependencies under true and false consensus. Triangles represent primary informants, squares represent secondary informants, and diamonds represent primary source "data". Emojis represent the reported claim - either supporting the claim (smiley face) or opposing (sad face). Panel A shows true consensus with independent data and primary sources and Panel C shows "true" consensus when primary sources share data. Panel B shows a false consensus in which seemingly converging reports originate from the same primary source.*

sources (Experiments 1A, 1B) or using a new scenario with more transparent causal relations between the data, primary informant, and secondary informant (Experiments 2, 3). These new experimental scenarios were designed to conform more closely to the *true consensus* structure illustrated in Figure 1A.

Experiments 1B, 2, and 3 used a modified informant consensus procedure that was arguably more sensitive to the detection of group-level differences in judgments in *true* and *false consensus* conditions. This study used a multi-stage approach wherein participants provided baseline confidence ratings after seeing one of four supporting essays and a single opposing essay. This allowed us to establish the degree to which confidence ratings change following repetition of claims from the same source or independent sources.



Given that the crucial YAK finding concerning an *illusion of consensus* was based on a null result in comparisons of confidence judgments in *true* and *false consensus* conditions, all of the current experiments used Bayesian statistical analyses to address the key research questions. Unlike the frequentist analyses used by YAK, such analyses allow us to quantify the strength of the evidence for the relevant null and alternative hypotheses (Lee & Wagenmakers, 2014).

## Experiment 1A

Our explanation for the *illusion of consensus* is that many people assume that the data used by the primary sources in the *true consensus* condition are not truly independent (i.e., their “independent” reports are derived from the same data/methods, as illustrated in Figure 1C). Experiment 1A aimed to address this issue through the addition of a novel “*enhanced true consensus*” condition where the instructions made it explicit that the multiple confirming reports from different primary sources were derived from independent data sets and methods (i.e., closer to the structure shown in Figure 1A).

## Method

### Participants

Three-hundred and twenty-three ( $M_{age} = 29.64$ ,  $SD_{age} = 10.13$ , females = 141, males = 175, other/undeclared = 5) participants completed the task via Prolific Academic (data from one participant who failed the attention check in the *true consensus* condition was discarded). The sample size was chosen to match the condition samples in YAK (Experiment 2). Participants were residents of the United States, United Kingdom, Canada, Ireland, Australia, or New Zealand (countries with at least an 80% English-speaking population), and were self-declared native English speakers. Participants in this and all subsequent experiments were paid £1.25 upon completion.

## Stimuli

The materials consisted of five fictitious student essays about a proposed tax policy in Sweden from YAK (the original materials used in YAK can be found at <https://osf.io/c3pzt/>). This scenario was chosen from the various YAK scenarios because a preliminary Bayesian statistical analysis of the original YAK data revealed that responses to this scenario provided the strongest null effect in comparisons of *true* and *false consensus* conditions (i.e., best evidence for an *illusion of consensus* - see <https://osf.io/nkg8m/> for analysis details)<sup>2</sup>. As per YAK, each essay contained six relevant facts, referred to in a different order. Although the wording of each essay differed, an attempt was made to minimize any differences between essays in substantive content other than facts relevant to the main claim. Essays ranged from 289 to 381 words. Four essays took a supporting stance and argued that the policy should be approved. One essay took an opposing stance and argued that the policy should not be approved. Participants were told that the students had been specifically instructed to cite the sources on which their conclusions were based. Each essay cited one primary source (e.g., the “Foundation for Economic Education”). The primary sources cited were all names of real economic foundations. Participants either read all five essays, with four pro and one con conclusions regarding the target claim (*true consensus*, *enhanced true consensus*, *false consensus*) or only two essays, one supporting and one opposing (*no consensus*).

## Procedure

Participants were randomly assigned to four conditions: a *true consensus*, *enhanced true consensus*, *false consensus*, or *no consensus* condition. The first three conditions were patterned after those used by YAK. The *enhanced true consensus* condition procedure was the same as *true consensus* except that, prior to presentation of the essays, there was an

---

<sup>2</sup> We are grateful to Yousif et al. (2019) for making their data and stimuli available in an online repository, and for providing further procedural details upon request.

extra screen of instructions. These instructions emphasized that the various economic foundations arrived at their conclusions using independent, but equally valid methods and data (*“Please assume that each of the sources that students cite in their essays have equal expert knowledge about the essay topic. The organisations cited in the essays are independent and may use different sources of data to arrive at their conclusions”*). The second sentence appeared in bold text.

After reading the instructions, participants completed a three-part comprehension check and were unable to proceed to the experiment until they had answered all three questions correctly. The main part of the experiment largely replicated the procedure in YAK Experiment 2 except that we highlighted each unique economic foundation with a different color. Participants in the *true consensus* conditions saw five sources highlighted in different colors and participants in the *false* and *no consensus* conditions saw two sources highlighted in different colors. Assignment of colors to sources was randomized across participants. This innovation was designed to improve participants’ memory for the primary sources reported in respective essays. Accurate source memory is a prerequisite for valid comparison of judgments in *true* and *false consensus* conditions.

Each participant read all five essays in each of the *true consensus* conditions (four supporting and one opposing). Each essay cited a single primary source; participants in the *true consensus* conditions heard from five distinct primary sources. Participants in the *false consensus* condition also read all five articles, but the four supporting essays cited the same source – thus, they only heard from two primary sources. Participants in the *no consensus* condition read two essays: one supporting and one opposing. Participants in all conditions were told that students had been asked to highlight each source using a different color to mark their essays. The same information was presented in each essay in all four conditions, and all that varied (other than superficial language) was the number of unique sources cited. Essays were presented to each participant sequentially in random order. The assignment of highlighting colors to specific economic foundation “sources” was randomized

across participants. Participants were told that had an unlimited amount of time to read each essay. They clicked the ‘Continue’ button, which appeared after a 30 second-delay, when they were ready to proceed to the next essay.

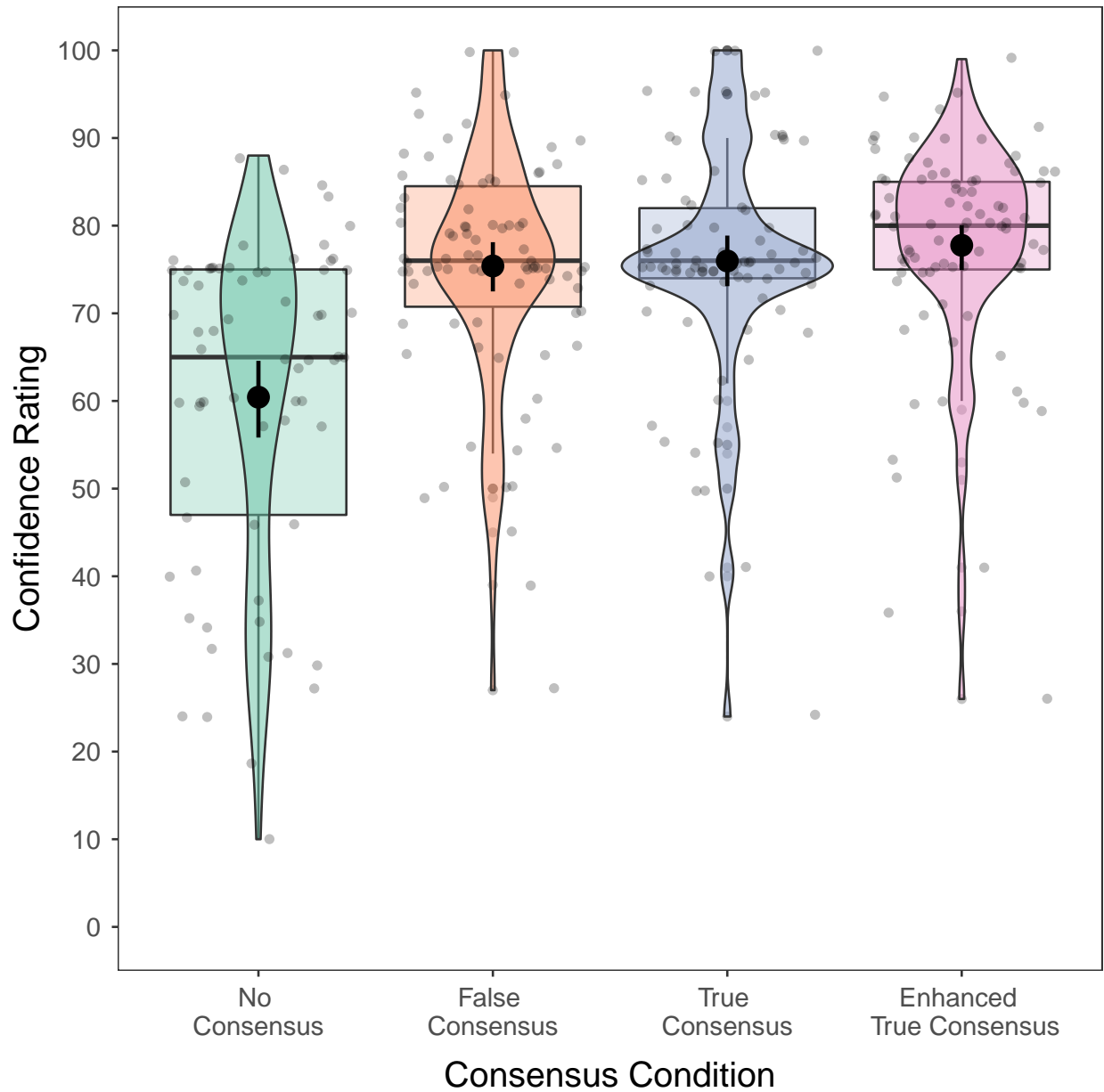
After reading all five essays, participants were presented with the following question: “Based on the arguments you saw, to what extent do you agree that the economy in question will continue to improve? Please rate your agreement on the scale below from 0 (strongly disagree) to 100 (strongly agree)”. Participants responded by clicking on a slider interspersed by five anchors. The numerical rating selected appeared below the slider, and participants were unable to proceed until they moved the slider thumb.

This was followed by a source memory test. On a separate page, participants saw a list of 10 possible sources (economic advisory bodies) and were asked to indicate which sources had been cited in the essays they read. The same source names were presented in all conditions. In the *true consensus* condition, five of these sources had been mentioned in the essays (“old” items) and five were new. In the *false* and *no consensus* conditions, two sources were old and eight were new.

The next phase involved checks on participants’ memory for the other key experimental details. A list of five nations was presented and participants were asked to identify the nation that was featured in the essays they read. Participants in all conditions also answered two post-test questions probing the number of supporting and opposing essays they had seen.

## Results and Discussion

Confidence ratings for the claim about tax policy are shown in Figure 2. Inferential analyses examining differences between group ratings in this and subsequent studies were conducted using the BayesFactor package with default priors (Morey et al., 2015). We followed the conventions suggested by Lee and Wagenmakers (2014), that a Bayes Factor ( $BF_{10}$  or  $BF_{01}$ ) between 3 and 10 represents “moderate” evidence for the alternative (or



**Figure 2**

*Violin plots show the distribution and probability density of confidence ratings in Experiment 1A. The violin plot is a symmetrical rotated kernel density plot and shows the density of the data at different values. Black points represent mean and 95% confidence interval of the mean. Boxes show median as a horizontal line and interquartile range. Grey dots show individual data points.*

null) hypotheses, a  $BF_{10}$  between 10-30 represents “strong” evidence, a  $BF_{10}$  between 30-100 represents “very strong evidence”, and a  $BF_{10}$  above 100 represents “extreme evidence”. We report  $BF_{10}$  when there is more support for the alternative than the null, and  $BF_{01}$  when the evidence favors the null hypothesis.

Differences between confidence ratings in each condition were examined using Bayesian t-tests with Cauchy default priors. We assessed the evidence for the hypothesis that confidence was greater under *true consensus* than *no consensus*. The Bayes factor analysis revealed extreme evidence for this difference,  $BF_{10} > 10,000$ . There was even stronger evidence of a difference between confidence in the *enhanced true consensus* and *no consensus* conditions;  $BF_{10} > 100,000$ . A comparison of *true* and *false consensus* conditions found moderate evidence for the null hypothesis,  $BF_{01} = 5.794$ . A comparison of *false consensus* and *no consensus* conditions found extreme evidence for the alternative,  $BF_{10} > 10,000$ .

Our key novel comparison involved the *enhanced true consensus* and *false consensus* conditions. Figure 2 shows that, as predicted, mean confidence was higher in the *enhanced true consensus* than the *false consensus* condition. Nevertheless, the Bayes factor analysis found that there was moderate evidence for the null when comparing confidence ratings in these conditions,  $BF_{01} = 3.191$ .

To check whether source memory accuracy affected the results we computed the number of correctly identified sources (hit rate) and number of incorrectly identified sources (false alarm rate) for each individual, replacing hit and false alarm rates of 0% or 100%, with values of 5% and 95%. Standard signal detection formulae (Macmillan & Creelman, 2004) were then applied to compute individual  $d'$  primes ( $d'$ ) for source memory. Participants with  $d'$  scores less than or equal to zero were then removed. We excluded 13 participants due to poor source identification (0 from the *false consensus* condition, 2 from the *no consensus* condition, 6 from *true consensus*, and 5 from the *enhanced true* condition). We also re-examined the data after excluding participants who substantially

miscounted the number of supporting/opposing essays.<sup>3</sup> A further 10 participants were excluded based on this criterion (3 from *false consensus*, 1 from *no consensus*, 4 from the *true consensus*, 2 from the *enhanced true* condition). These exclusions did not change the key results. Full details of the analyses can be found at <https://osf.io/nkg8m/>.

This experiment replicated the *illusion of consensus* results from YAK (Experiment 2). Multiple reports in favor of a claim from different sources (*true consensus*) or the same source (*false consensus*) increased confidence in the claim relative to the *no consensus* condition. Crucially, the Bayesian analyses found moderate evidence for the null effect when comparing *true* and *false consensus* conditions. The novel *enhanced true consensus* condition did result in a numerically higher mean confidence rating than either the *false* or *true consensus* conditions, but there was no evidence from the Bayes factor analysis that this was a robust difference.

Figure 2 however, reveals a high level of variability in confidence ratings between participants within each condition. For example, in the *true consensus* condition, Figure 2 shows a range of around 80 points between the highest and lowest ratings, with several participants giving ratings close to or below the mid-point of the confidence scale. This was despite all these participants having read the same four supporting essays and one opposing essay. This variability may, in part, reflect differences in the way that individuals within a given condition used the confidence scale. Given that the target claim about tax policy was related to a domain likely to be unfamiliar to most participants, and that there was no way of objectively verifying the claim, it is understandable that some participants

---

<sup>3</sup> The number of supporting essays was four in the *true* and *false consensus* conditions and one in the *no consensus* condition and the number of opposing essays was one in all conditions. We subtracted participant's estimates for the number of opposing essays from the supporting essays and excluded anyone with an error greater than two. Participants in the *true* and *false consensus* conditions who answered correctly had scores of 3 and participants in the *no consensus* condition had scores of 0. Thus, anyone in the *true* and *false consensus* conditions with a score less than 1 or greater 5 was excluded and anyone with a score greater than 2 or less than -2 was excluded from the *no consensus* condition.

may have given conservative ratings, even if they were somewhat persuaded by the consensual evidence. This high level of variability in the confidence measures likely reduced the sensitivity of our test of differences between the *false* and enhanced *true consensus* condition. This issue was addressed in Experiment 1B.

## Experiment 1B

One way of addressing the problem of individual variability in the use of the confidence scale is to introduce “baseline” confidence ratings after presentation of one pro and one con report, but before exposure to additional reports that represent *true* or *false consensus*. The key research question then becomes whether subsequent reports that reflect *true consensus* led to larger shifts in ratings from this baseline than repeated reports in the *false consensus* condition. Hence, in Experiment 1B confidence in the target claim was measured after all participants had viewed one supporting and one opposing report from different sources and again after all reports were presented. As well as allowing for more control over variability in scale use, this procedure may offer a more sensitive test of the effects of *true* and *false consensus* because participants are explicitly asked to consider the *change* in their beliefs about the claim because of further information from independent sources or the same source. Note that, in this case, all participants provided what were in effect “*no consensus*” ratings in the first stage of the procedure. Hence there was no need to run a separate *no consensus* condition.

## Method

### Participants

Four-hundred and sixty ( $M_{age} = 33.07$ ,  $SD = 12.24$ , females = 228, males = 225, other/undeclared = 6) participants completed the experiment via Prolific Academic (data from 10 participants who failed the attention check in the *true consensus* condition was discarded). We used the same selection criteria, and aimed to recruit the same number of



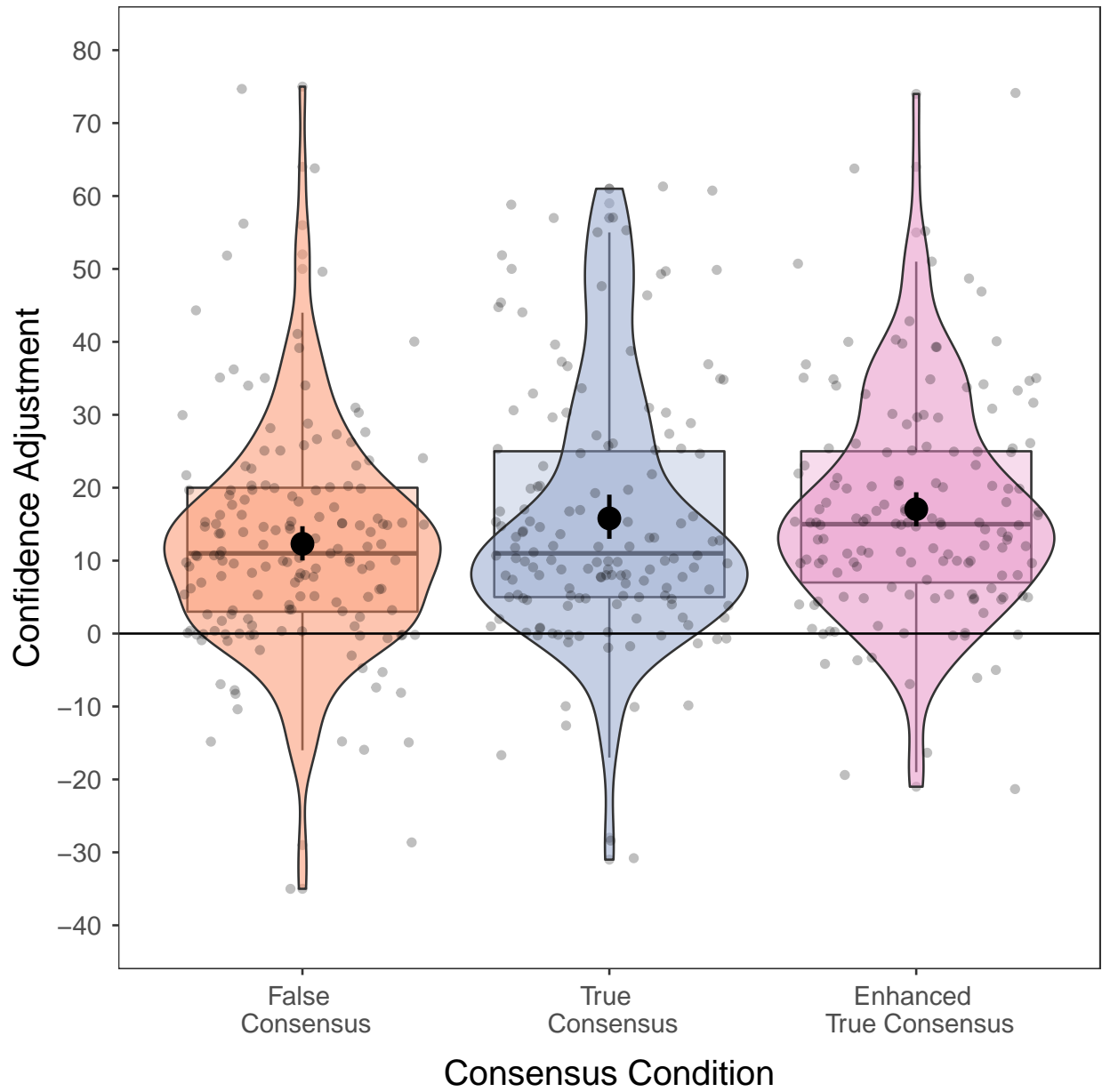
participants, as Experiment 1A.

## **Stimuli**

Materials were identical to Experiment 1A.

## **Procedure**

Participants were randomly allocated to one of three consensus conditions. The procedure was like Experiment 1A except that participants in Experiment 1B provided two confidence ratings. The first rating was made after reading one supporting and one opposing essay. A second rating was made after presentation of the remaining three supporting essays (either citing the same or three different primary sources). The supporting essay that participants saw before providing their first confidence rating was randomly selected from the full set of four supporting essays. The presentation order of essays was randomized (before and after initial confidence ratings), as were the primary sources' names, and highlighting colors.



**Figure 3**

*Violin plots show the distribution and probability density of adjustment in confidence ratings from the baseline to the final rating in Experiment 1B. Black points represent mean and 95% confidence interval of the mean. Boxes show the median as horizontal grey line and interquartile range. Grey dots show individual data points. The horizontal black line indicates no change in confidence from the baseline.*

## Results and Discussion

In Experiment 1B, we were interested in the extent to which participants adjusted their confidence ratings after seeing a *true* (independent) or *false* (dependent) consensus. If participants are sensitive to *false consensus*, we would expect to see a greater increase in confidence in the *enhanced true consensus* condition than the *false consensus* condition. We subtracted the first confidence rating from the second to compute confidence adjustment scores, which are shown in Figure 3.

We performed Bayesian t-tests (with a Cauchy default prior) to compare confidence adjustments between the experimental conditions. A comparison of the *true* and *false consensus* conditions found that although there was a greater increase in confidence in the *true consensus* condition, the evidence in favor of this difference was anecdotal,  $BF_{10} = 1.27$ . Notably, as predicted, a comparison of adjustments in the *enhanced true* and *false consensus* conditions revealed moderate evidence in favor of a larger confidence increase based on *enhanced true consensus*,  $BF_{10} = 3.88$ .

We again checked whether source memory accuracy affected the results (excluding participants with zero or negative  $d'$  scores: 1 from *false consensus*, 2 from *true consensus*, 4 from *true consensus enhanced*). We also compared the experimental conditions after excluding participants who miscounted the number of supporting/opposing essays (a further 11 from *false consensus*, 2 from *true consensus*, and 7 from *true consensus enhanced*). Neither re-analysis substantially changed the key results (see <https://osf.io/nkg8m/> for full details).

We also examined whether people entirely discounted the repetition of information in the case of *false consensus* - that is, whether final confidence ratings after hearing repetitions of reports supporting the claim from the same source, deviated significantly from the baseline ratings. We performed a one-sample Bayesian t-test to examine the evidence that confidence adjustment in the *false consensus* condition was not equal to zero. There was moderate in favor of the alternative,  $BF_{10} = 5.47$ . This suggests that people do

attach some weight to information repeated from the same source.

Experiment 1B again found some evidence for the *illusion of consensus* in that people did not always attach more weight to information provided under *true consensus* compared to *false consensus*. Crucially, however, people were more persuaded about the truth of a claim in the *enhanced true* than the *false consensus* condition. This result supports our central hypothesis - that when people believe that a set of converging reports arise from truly independent data sources, they will assign greater epistemic weight to these reports than to repeated reports from a single source.

## Experiment 2

Experiment 1B showed that reducing ambiguity about the independence between the primary sources in a *true consensus* condition can boost sensitivity to the epistemic value of *true* over *false consensus*. Experiment 2 examined a similar prediction using a different approach. Instead of giving an explicit instruction about source independence we developed a task scenario where the causal relations between the original data, primary and secondary sources were likely to be more familiar and transparent to participants.

Political election polling, whereby the popularity and predicted voting results for candidates are surveyed and reported before an election is held, are a ubiquitous and growing phenomenon in contemporary democracies (Hillygus, 2011). In most democratic polities several such polls are available for a given election, with poll producers often advertising the novelty and robustness of their polling methods relative to their competitors. Given that most participants could be assumed to have some familiarity with such polling processes, we used a polling scenario to examine sensitivity to the impact of *true* and *false consensus*. Hence, in these scenarios, the “data” were the results of polls carried out by separate polling companies (primary sources), which were then reported as social media ‘Tweets’ by news agencies (secondary sources). Predicted voting trends favoring one of two possible candidates were reported by multiple, independent polling

companies (*true consensus*) or repeatedly reported by a single company (*false consensus*).<sup>4</sup>

Because we expected that the causal relationships between data, primary and secondary sources would be more transparent than in Experiments 1A-B, we did not include an *enhanced true consensus* condition. This avoids a possible limitation of the previous studies. In those studies, the extra instructions explaining data independence may have been seen as a cue that the experimenters expected participants to pay extra attention to this evidence (regardless of the details about source independence). This possible task demand factor was eliminated from the current study.

In designing the scenarios for this study, we also aimed to reduce the amount of novel, extraneous information that was added to each report. Although the tax policy reports in Experiment 1A-B and YAK controlled for the number of key facts presented, the way that these facts were described (e.g., the benefits of the tax policy are quantified in monetary terms in one essay and as increases in employment in another) and the inclusion of other extraneous details (e.g., exactly who will benefit from the policy) varied between scenarios, including those repeated in the *false consensus* condition. Adding details that have no evidentiary value to a report can nevertheless increase confidence in that report (Heit & Rotello, 2012; Weisberg et al., 2008). This could have led to a spurious inflation of confidence in the repeated *false consensus* reports. To address this issue, in Experiment 2 we developed a scenario with minimal extraneous text – the information relevant to the target claim about voting references was presented in the same linguistic frame in all reports, both independent and repeated.

---

<sup>4</sup> This experiment was completed before the 2020 US election



Figure 4

*Example stimuli and experimental conditions in Experiment 2. The top panel shows the **true consensus** condition in which four unique polling companies are cited. The bottom panel shows the **false consensus** condition in which the same polling company is cited four times.*

## Method

### Participants

One-hundred and ninety-five ( $M_{age} = 31.02$ ,  $SD = 11.65$ , females = 65, males = 125, other = 5) participants completed the experiment via Prolific Academic (data from 11 participants who failed the attention check and an additional participant with a missing response to the attention check question were removed prior to analysis). The selection criteria were identical to the previous experiments.<sup>5</sup>

### Stimuli

Figure 4 shows example materials from Experiment 2. Materials consisted of a series of Tweets (brief messages posted on a social media platform) from fictitious news outlets. The ‘Tweets’ constituted polling predictions from either the same or different polling companies. Each polling company was labeled with a different Greek letter (Gamma, Delta, Kappa, Sigma, and Theta) and predicted whether Candidate X or Candidate Y would receive more votes in an upcoming election. Participants in the *true consensus* condition saw predictions from five different polling companies, whereas participants in the *false consensus* condition only saw predictions from two different polling companies, with one prediction repeated four times.

### Procedure

Participants were randomly allocated to either the *true* or *false consensus* condition. They were asked to imagine an upcoming election in a foreign democratic country with which they were unfamiliar. They were told that they would see ‘Tweets’ from news outlets posted about polling results for projected election results. Each polling company was

---

<sup>5</sup> There were more attention-check based exclusions in Experiment 2 than the previous experiments. The attention check format was the same but the response options were arguably less easy to discriminate (see Procedure below)

labelled with a unique Greek letter and highlighted in a separate color. Participants were told that each polling company had conducted an independent poll using their own expert methods. Each of the polling organizations was described as equally reputable and having comparable track records in accurately predicting elections.

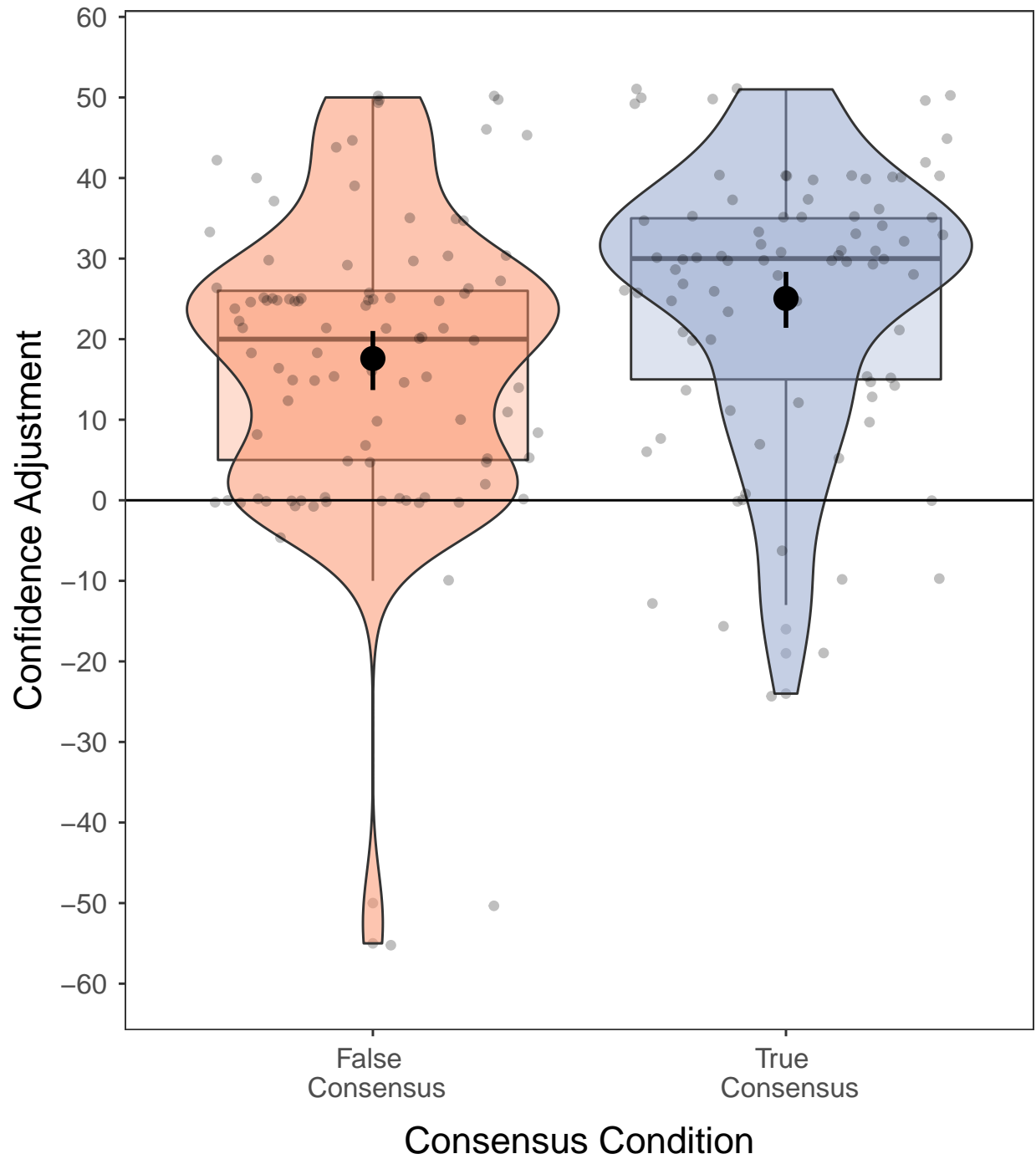
Participants completed a three-part instruction check and could not proceed to the main study until all three questions had been answered correctly. Following this, participants saw one Tweet predicting that Candidate X would receive more votes in the upcoming election and one Tweet predicting that Candidate Y would receive more votes in the upcoming election, with order randomized across participants. Participants then answered the question: *“Based on the polls you saw, to what extent do you agree that Candidate X/Y will win the upcoming election? Please rate your agreement on the scale below from 0 (strongly disagree) to 100 (strongly agree)”*. Immediately after providing an initial confidence rating, participants were told they would now see more Tweets about the election and reiterated that the poll company name would be highlighted with the same color used for the same polls and different colors used for different polls. We randomized the polling company names and whether Candidate X or Y was predicted to win in the majority of Tweets across participants.

After reading all five Tweets, participants answered the same question about which candidate they thought would win in the upcoming election. Participants responded by clicking on a slider interspersed by five anchors to indicate their confidence that the candidate with the majority will receive more votes in the upcoming election.

On a separate page, participants saw a list of 10 poll company names with Greek letter labels and were asked to indicate which had been cited in the Tweets to test their source memory. Following this, we asked participants which labels we used to refer to the candidates in the election (options: X & Y, X, Y & Z or A & B). Participants in both conditions answered two post-test questions probing the number of Tweets favoring Candidate X/Y they had seen. Following this, participants were asked to justify their



second confidence rating. Participants in the *true consensus* condition answered the question “*The Tweets you saw all cited different polls. How did this influence your judgment about which candidate would get more votes?*”. Participants in the *false consensus* condition answered a similar question but the question specifically mentioned the poll they had seen repeated.

**Figure 5**

*Violin plots show the distribution and probability density of confidence ratings in Experiment 2. Black points represent mean and 95% confidence interval of the mean. Boxes show median as a horizontal grey line and interquartile range. Grey dots show individual data points. The horizontal black line indicates no change in confidence from the baseline.*

## Results and Discussion

Figure 5 shows the mean confidence adjustment scores in Experiment 2. As with previous experiments, we performed Bayesian t-tests (with a Cauchy default prior) to examine the strength of the evidence for the alternative hypothesis that there would be a greater increase in confidence in the *true* than the *false consensus* condition.

There was moderate evidence for the alternative for a greater increase in confidence from baseline to final confidence ratings in the *true* than the *false consensus* conditions,  $BF_{10} = 6.297$ .

We again checked whether source memory accuracy or memory for the number of Tweets favoring each candidate affected the results, but neither analysis substantially changed the key result. Hence, our main analysis showed that people were able to discriminate between the epistemic warrant of *true* and *false consensus* (see <https://osf.io/nkg8m/> for full details).

We also examined whether those in the *false consensus* condition entirely discounted the repetition of information from the same source. A one-sample Bayesian t-test examined the evidence that confidence adjustment in the *false consensus* condition was not equal to zero. There was very strong evidence in favor of the alternative,  $BF_{10} > 10,000$ . As in Experiment 1B, this suggests that people do attach some weight to information repeated from the same source.

To summarize, there was a larger increase in confidence in claims about which candidate would receive more election votes in the *true consensus* condition than the *false consensus* condition. Experiment 2 shows clear evidence of sensitivity to the difference between *true* and *false consensus* when the scenario provides a clearer mapping of the independence/dependence relations between the data and primary sources. In this study we found no *illusion of consensus*. Within the *false consensus* condition, however, participants did not entirely discount the repetition of information from the same primary source.

*Confidence justifications*

We carried out an exploratory examination of participants' written justifications for their confidence ratings. Given that the justification question differed between *true* and *false consensus* conditions, responses from each condition were coded separately. There were four response categories but we were primarily interested in the proportion of responses that fell into the first two categories <sup>6</sup>. For the *false consensus* condition the two main categories of interest were responses that indicated:

1. Repetition of the same poll by multiple news outlets had a positive influence on their judgment (e.g. makes it more credible/reliable).
2. Repetition of the same poll had a negative influence on their judgment (e.g., redundant).

For the *true consensus* condition these categories were:

1. Different polls being cited by different news outlets had a positive influence on their judgment.
2. Different polls being cited had a negative impact on their judgment (e.g., it made them less sure, there were more sources to integrate).

Justifications were first coded by two independent coders and discrepancies between coders were resolved by a third coder. In the *true consensus* condition 55% of the responses indicated that different news outlets citing different polls had a positive influence on their confidence judgments. For example, one participant said: "*Multiple polls indicating the same outcome suggest a higher likelihood of one candidate winning although polling data can never predict the outcome with certainty*". In the *false consensus* condition, 47% of

---

<sup>6</sup> The final two categories related to a) responses indicating that only the number of Tweets and/or news outlets were important regardless of source, and b) "other"

participants indicated that same poll being cited had a positive influence on their confidence judgment (e.g., “*Since all the news outlets were citing Delta they must have had a good reason to think Delta’s poll would be predictive of the results of the election*”). Very few in the *true consensus* condition indicated that viewing multiple independent polls had a negative effect on confidence (2%). A minority in the *false consensus* condition indicated that repetition of the same poll had a negative impact on their judgment (17%; e.g., “*All tweets referenced the same poll, so I only had one source of data*”). These confidence justifications indicated that one reason that participants give weight to a *false consensus* is that repetitions of a claim by different secondary sources bolsters the claim’s credibility.

### Experiment 3

Experiment 2 showed that sensitivity to the epistemic value of *true* over *false consensus* can be boosted when the scenario provides a clearer mapping of the independence/dependence relations between the data and primary sources. One notable difference between YAK’s original experiments (and Experiments 1A and 1B), and Experiment 2 is that, unlike student essays or news articles, the Tweets about poll results may not be treated as endorsements of a particular view or claim. In the news Tweets used in Experiment 2 there was no ‘author’ trying to justify an argument or offering commentary on the poll findings. Indeed, people often Tweet about things they do not necessarily endorse (e.g., a left-leaning media outlet may tweet the results of a right-leaning poll).

It may be that the *illusion of consensus* is stronger when the secondary source is perceived as *endorsing* the primary source’s claim, as in the student essays and news commentaries in YAK.<sup>7</sup> The goal of the current study therefore was to examine whether our previous finding of a reduction in the *illusion of consensus* was affected by whether the information provided in *true* and *false consensus* conditions was endorsed by the source who shared it.

---

<sup>7</sup> This idea was suggested to us by a reviewer (Sami Yousif).

Experiment 3 examined this issue by comparing confidence ratings in *true* and *false consensus* conditions under two levels of endorsement of the shared information. The *no endorsement* conditions replicated the experimental scenarios used in Experiment 2 (news reports of election poll results). The new *endorsement* condition had a similar structure except that the reports of poll results were made by individual tweeters and contained clear evidence of a favorable endorsement of the result (via positive emojis). We pre-registered the method, directional hypotheses, and analysis plan (including planned analyses, data stopping rule, and exclusion criteria) before data collection; the pre-registration details can be found at <https://osf.io/fk76w>.

## Participants

Three-hundred and sixty-five ( $M_{age} = 35.48$ ,  $SD = 13.58$ , females = 194, males = 155, other = 16) participants completed the experiment via Prolific Academic (data from 18 participants who failed the attention check were removed prior to analysis). The selection criteria were identical to the previous experiments.

**Figure 6**

*Example stimuli for the **endorsement** conditions in Experiment 3. The top panel shows the **true consensus endorsement** condition in which four unique polling companies are cited. The bottom panel shows the **false consensus endorsement** condition in which the same polling company is cited four times.*

## Stimuli

Materials in the *no endorsement* conditions were identical to Experiment 2. Figure 6 shows the new materials for the *endorsement* conditions. These were similar to the fictitious Tweets from the previous study, except that the secondary source was an individual rather than a news outlet (see Figure 6 for examples). Each individual ‘Tweeter’ also used three emojis (randomly selected from a set of six positive emojis) to demonstrate their endorsement of the poll being Tweeted.<sup>8</sup>

## Procedure

Participants were randomly allocated to one of four conditions: the *true consensus - endorsement*, *false consensus - endorsement*, *true consensus - no endorsement*, or *false consensus - no endorsement*. The procedure was almost identical to Experiment 2 except that participants in the *endorsement* conditions saw references to people/persons instead of news outlets and received an additional instruction that the ‘Tweeters’ also showed what they think about the polls using positive emojis.

## Results and Discussion

Figure 7 shows the mean confidence adjustment scores in Experiment 3. To quantify the relative effects of *consensus* and *endorsement*, we conducted a 2 (consensus) x 2 (endorsement) factorial Bayesian analysis of variance. We obtained Bayes factors representing the relative odds of the data occurring under models including each main effect, and the main effects and interaction, as compared against an intercept-only model. This analysis found the strongest evidence for the model including the main effects of *consensus* (*true consensus* vs. *false consensus*) and *endorsement* (*endorsement* vs. *no endorsement*), but no interaction ( $BF_{10} = 249.97$ ). The model including only the two main

---

<sup>8</sup> We also changed the polling company ‘Delta’ to ‘Eta’ due to the frequency with which the former letter has been used recently in the media.



effects was preferred over a saturated model including main effects and the interaction ( $BF_{10} = 5.73$ ). Examination of Figure 7 shows that, across *endorsement* conditions, those in the *true consensus* condition showed a larger increase in confidence ratings than those in the *false consensus* condition. Across *consensus* conditions, confidence ratings were generally higher in the *no endorsement* than the *endorsement* condition.

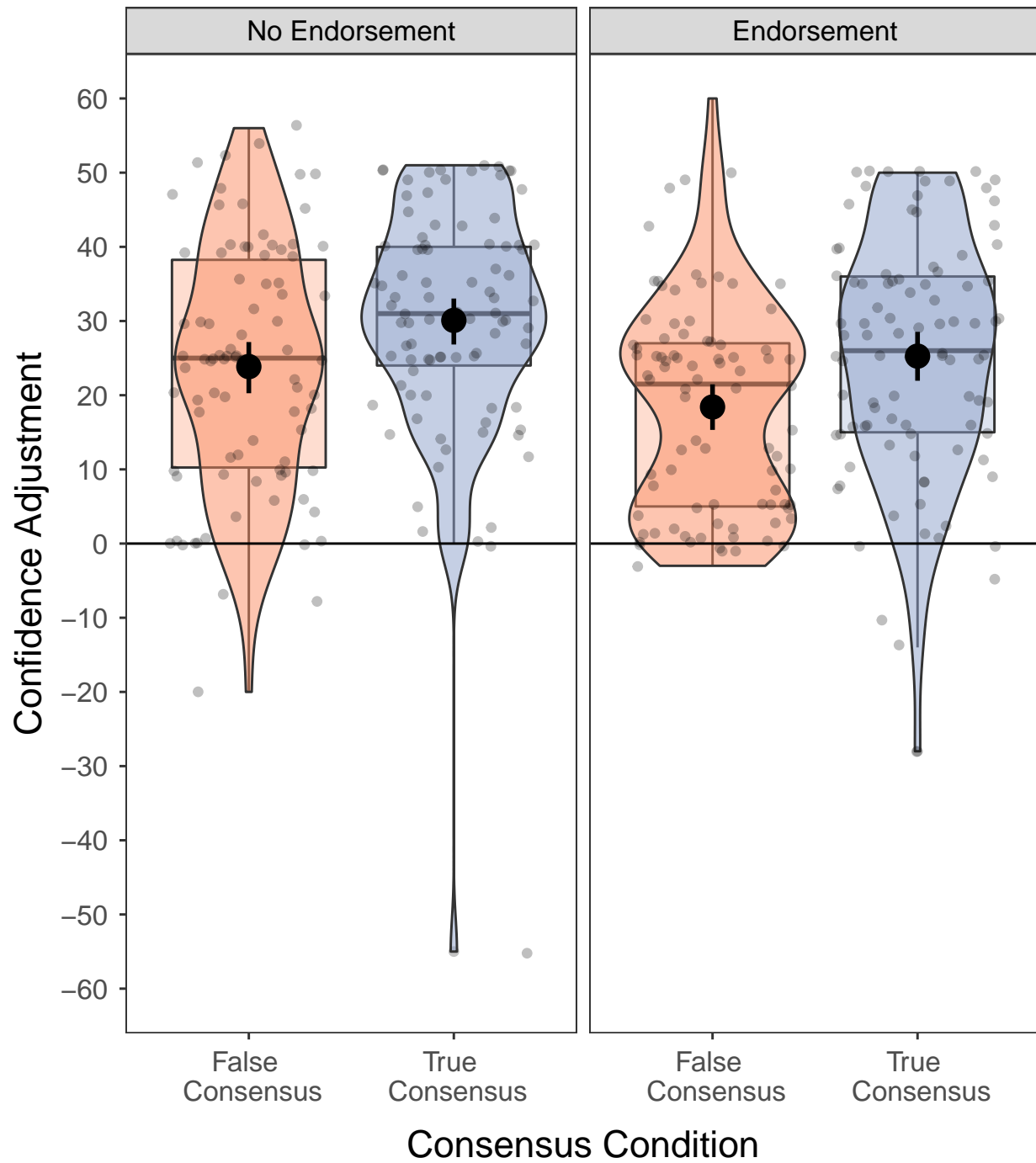
As with previous experiments, we checked whether source memory accuracy affected or memory for the number of Tweets favoring each candidate affected the results, but neither analysis substantially changed the key results (see <https://osf.io/nkg8m/> for full details). The only notable change was that when participants with poor memory for number of Tweets were excluded, there was weaker evidence for an effect of endorsement on confidence (the preferred model contained only a main effect for consensus,  $BF_{10} = 40.015$ ).

We also examined whether those in the *false consensus* conditions entirely discounted the repetition of information from the same source. We examined the evidence that confidence adjustment in the *false consensus* condition was not equal to zero with separate one-sample t-tests for the *endorsement* and *no endorsement* conditions. There was moderate evidence that confidence adjustment differed from zero in the *endorsement* condition ( $BF_{10} = 4.009$ ) and anecdotal evidence in the *no endorsement* condition ( $BF_{10} = 2.624$ ). These results again suggest that people do attach some weight to information repeated from the same source.

In sum, across *endorsement* conditions, we found a greater increase in confidence in tweeted claims in the *true consensus* conditions than the *false consensus* conditions. Replicating the key finding of Experiment 2, we found evidence of sensitivity to the difference between *true* and *false consensus* when the scenario provides a clear mapping of the independence/dependence relations.

Larger adjustments in confidence were made when there was *no endorsement* of the information being shared. It appears that knowing that an individual sharing a Tweet endorses the tweet content led to a general reduction in confidence in the tweeted claim.

Crucially, however, this effect of *endorsement* did not interact with the obtained difference in confidence between *true* and *false consensus* conditions.



**Figure 7**

Violin plots show the distribution and probability density of confidence ratings in Experiment 3. Black points represent mean and 95% confidence interval of the mean. Boxes show median as a horizontal grey line and interquartile range. Grey dots show individual data points. The horizontal black line indicates no change in confidence from the baseline.

## General Discussion

The current studies re-examined the *illusion of consensus* – in which people give similar epistemic weight to reports provided by a majority of separate, independent sources and the same number of reports repeated from a single source (Yousif et al., 2019). We investigated a novel theoretical account of this illusion - arguing that it arises because of ambiguity in the causal relations between the sources of evidence in the *true consensus* condition (i.e., a belief that reports in this condition were not developed independently of one another). An initial test of this hypothesis (Experiment 1A) failed to find support, but did replicate the *illusion of consensus* in *true* and *false consensus* conditions that were similar to those used by YAK. The hypothesis concerning transparency of independence between sources was, however, supported in subsequent experiments where we measured change in participants' confidence in a target claim as they received additional reports from independent sources or the same source. In Experiment 1B, adding an explicit statement about the independence between reports in the *true consensus* condition led participants to give more weight to this evidence than the repeated reports in the *false consensus* condition. Experiments 2 and 3 employed a new scenario which further highlighted the independence between reports in the *true consensus* condition. More weight was subsequently given to reports in this *true consensus* condition than to a *false consensus* condition – there was no *illusion of consensus*.

The results of Experiment 1A, and those of the *true* and *false consensus* conditions in Experiment 1B, show that people often give similar weight to reports derived from *true* and *false consensus*. Crucially, however the results of Experiments 1B, 2 and 3 also show that people have the *capacity* to distinguish between *true* and *false consensus* when the relationship between primary source and the data used to develop an argument is made explicit. Our results suggest that a key reason for previous failures to find a difference in the epistemic weight given to *true* and *false consensus* (Foster et al., 2012; Sulik et al., 2020; Yousif et al., 2019), was ambiguity about the independence of the information

contained in *true consensus* reports.

We believe that this explanation of the *illusion of consensus* applies to many examples of belief formation outside the lab. One reason that people may adopt conspiratorial beliefs is that they incorrectly perceive causal dependencies amongst independent sources. For instance, assuming that climate scientists working independently are members of a ‘global cabal’ with a common motivation to deceive the public and control the climate debate (cf. Rabb, Fernbach & Sloman, 2019).

On the positive side however, our results suggest that if people can be convinced about the independence of primary data sources, they will accord more weight to *true* than *false consensus*. This work suggests that an important avenue for future work is to investigate peoples’ beliefs about source dependence in a given scenario and examine the most effective ways of communicating source independence, when it exists.

Theoretically, the current results are important because they suggest that when people are provided with transparent information about the causal in(dependence) of information sources, they can follow the prescription of normative models of inference (e.g., Whalen et al., 2018), at least in a qualitative sense. That is, they attach more weight to consensus among informants reporting on data obtained from independent sources than dependent sources. In this regard, our findings mirror demonstrations of sensitivity to dependence between informant evidence in tasks like probabilistic judgment (e.g., Whalen et al., 2018), and advice-giving (e.g., Yaniv et al., 2009).

Our theoretical approach has some similarity to other cases where apparently non-normative or irrational judgments are due to people having incomplete or erroneous causal beliefs. Krynski and Tenenbaum (2007) for example, suggested that the apparent neglect of base-rates in probabilistic reasoning tasks may be the result of an impoverished causal representation of the problem. When the causal relations between problem elements were clarified, the accuracy of probabilistic reasoning increased (Hayes et al., 2016; Hayes et al., 2014). Tools that help people to understand causal structure can also improve

performance on complex probabilistic reasoning problems, involving multiple dependencies (Cruz et al., 2020).

Our results are consistent with Mercier and Miton’s (2019) findings that give less weight to dependent consensus which is based on direct repetition of claims made by other individuals. What is less clear is whether our results support Mercier and Miton’s more general claim about the primacy of “evolutionarily valid” cues to dependency. They suggest that people will be more sensitive to cues to dependence that are presumed to have a long evolutionary history (e.g., knowing that individuals discussed the evidence) than those that are more recent cultural innovations (e.g., providing the statistical correlation between reports; Maine, 1990). It is not clear where the cues to dependency used in the current studies, such as reliance on a common expert source, fit within Mercier and Miton’s framework. It is also unclear how the effects of the evolutionary validity of cues can be differentiated from the effects of familiarity or frequency of cue use within a culture. We also note that non-experts are capable of understanding “evolutionarily recent” statistical concepts such as correlation and regression, when these are presented in examples drawn from a familiar domain (Nisbett et al., 1983), and that the use of such cues can be readily improved through training (Fong et al., 1986; Sedlmeier, 1999).

The present results also have implications for YAK’s claim that people are sensitive to *false consensus* when the claim in question is *knowable*. In their Experiment 5, YAK found that participants substantially discounted *false consensus* when testimony was based on a directly perceived event. Here, YAK had participants read an article involving a bear sighting in which the same eyewitness or multiple eyewitnesses testified that they saw the bear invade a schoolyard. YAK argued that the claim in this scenario is *knowable* in the sense that no special expertise would be required to establish the truth of the claim. By contrast, in all of YAK’s other experiments, most participants would have to rely on advice from others to evaluate claims about the impact of complex economic policies. Our results suggest a different interpretation of the YAK results in the bear-scenario. Successful

discrimination between *true* and *false consensus* in the bear scenario may have been due to the relatively familiar and transparent causal relationships between the “data” (e.g., sighting of the bear) and the successive informant reports. In such a scenario the independence between alternative eyewitnesses (assuming they had not communicated with each other) would be relatively easy to grasp.

### Other factors affecting judgments of consensus

Although we found that people can distinguish between the epistemic value of *true* and *false consensus*, there was strong evidence in each study that people did not entirely discount information repeated from a single source. In Experiment 1A, confidence ratings in the *false consensus* condition were well above those in the *no consensus* condition - which was equivalent in the amount of novel information provided about the target claim. Likewise in Experiments 1B, 2 and 3, participants in the *false consensus* condition showed a significant increase in confidence in a claim after repetition of support for the claim from a single source.

There are at least two possible explanations for such effects of claim repetition, which are not mutually exclusive. The first is that repetition increases the familiarity of the claim - enhancing its memory representation, and that this is sufficient to increase confidence in the claim (Begg et al., 1992; Weaver et al., 2007). Note, however, that this explanation does not apply to comparisons between *true* and *false consensus* conditions, where claim frequency is matched.

The second reason why people might attach more weight to a claim as it is repeated is that participants make inferences about *why* the repetition occurred (Colantonio et al., 2021). For example, in the *false consensus* conditions, participants may have inferred that multiple secondary sources cited a single primary source with good reason. The secondary sources (e.g., news outlets in Experiment 2) may have selected that primary source from available alternatives because they believed it was the most reputable, credible or

authoritative (e.g., Estlund, 1994). This would serve to inflate confidence in the claim under false consensus.

We found some evidence for this mechanism in the justifications that people provided for confidence ratings in Experiment 2. A substantial proportion of those in the *false consensus* condition believed that repetition of the information from a single source implied that the source was especially credible. This is likely to be an additional factor contributing to the *illusion of consensus*. The current results, however, show that this effect can be offset by strengthening belief in the independence of the primary sources in true consensus conditions.

The current work was concerned with establishing the conditions under which people do or do not discriminate between the epistemic warrant of *true* and *false consensus*. Our overarching assumption was that *true consensus* should be given more weight when evaluating informant claims. It is important to note, however, that there are some circumstances where repeated observations from a single source offer as much if not more information than reports from multiple, independent sources. When a single source has greater domain expertise or more reliable methods for determining the ground truth than other sources (e.g., a scientist who has developed a well-controlled paradigm for observing a phenomenon) then it is rational to give more weight to repeated claims from this source than from independent but less expert sources. Moreover, Pilditch et al. (2020) have shown that when the reports of successive informants are contradictory rather than corroborative, dependency between informants can have advantages over having independent informants (e.g., where a less expert informant who has reached a particular conclusion, revises their belief after receiving a dissenting report from a more expert informant).

## Conclusions

The current studies support the view that an *illusion of consensus* can arise when people are unsure about the independence of the primary sources on which informant



claims are based. Our studies re-confirmed that people are often susceptible to this illusion. Crucially, however, we have shown that when relations between information sources are made transparent, people do assign more epistemic weight to true than to false consensus. Our findings therefore advance our understanding of cognitive mechanisms that drive the *illusion of consensus*. They also point to ways of overcoming this illusion.

## References

- Aboody, R., Yousif, S. R., Sheskin, M., & Keil, F. (2021). Says who? children consider informants' sources when deciding whom to believe.
- Anderson, L. R., & Holt, C. A. (1997). Information cascades in the laboratory. *The American economic review*, 847–862.
- Begg, I. M., Anas, A., & Farinacci, S. (1992). Dissociation of processes in belief: Source recollection, statement familiarity, and the illusion of truth. *Journal of Experimental Psychology: General*, 121(4), 446.
- Bovens, L., Hartmann, S. et al. (2003). *Bayesian epistemology*. Oxford University Press on Demand.
- Colantonio, J., Durkin, K., Caglar, L. R., Shafto, P., & Bonawitz, E. (2021). The intentional selection assumption. *Frontiers in psychology*, 4664.
- Cruz, N., Connor Desai, S., Dewitt, S., Hahn, U., Lagnado, D., Liefgreen, A., Phillips, K., Pilditch, T., & Tešić, M. (2020). Widening access to bayesian problem solving. *Frontiers in psychology*, 11, 660.
- Dechêne, A., Stahl, C., Hansen, J., & Wänke, M. (2010). The truth about the truth: A meta-analytic review of the truth effect. *Personality and Social Psychology Review*, 14(2), 238–257.
- Dietrich, F., & Spiekermann, K. (2013). Independent opinions? on the causal foundations of belief formation and jury theorems. *Mind*, 122(487), 655–685.
- DiFonzo, N., Beckstead, J. W., Stupak, N., & Walders, K. (2016). Validity judgments of rumors heard multiple times: The shape of the truth effect. *Social Influence*, 11(1), 22–39.
- Estlund, D. M. (1994). Opinion leaders, independence, and condorcet's jury theorem. *Theory and Decision*, 36(2), 131–162.

- Fazio, L. K., Brashier, N. M., Payne, B. K., & Marsh, E. J. (2015). Knowledge does not protect against illusory truth. *Journal of Experimental Psychology: General*, *144*(5), 993.
- Fong, G. T., Krantz, D. H., & Nisbett, R. E. (1986). The effects of statistical training on thinking about everyday problems. *Cognitive psychology*, *18*(3), 253–292.
- Foster, J. L., Huthwaite, T., Yesberg, J. A., Garry, M., & Loftus, E. F. (2012). Repetition, not number of sources, increases both susceptibility to misinformation and confidence in the accuracy of eyewitnesses. *Acta psychologica*, *139*(2), 320–326.
- Hahn, U., von Sydow, M., & Merdes, C. (2019). How communication can make voters choose less well. *Topics in Cognitive Science*, *11*(1), 194–206.
- Harvey, J. A., Van Den Berg, D., Ellers, J., Kampen, R., Crowther, T. W., Roessingh, P., Verheggen, B., Nuijten, R. J., Post, E., Lewandowsky, S., et al. (2018). Internet blogs, polar bears, and climate-change denial by proxy. *BioScience*, *68*(4), 281–287.
- Hayes, B. K., Hawkins, G. E., & Newell, B. R. (2016). Consider the alternative: The effects of causal knowledge on representing and using alternative hypotheses in judgments under uncertainty. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(5), 723.
- Hayes, B. K., Hawkins, G. E., Newell, B. R., Pasqualino, M., & Rehder, B. (2014). The role of causal models in multiple judgments under uncertainty. *Cognition*, *133*(3), 611–620.
- Heit, E., & Rotello, C. M. (2012). The pervasive effects of argument length on inductive reasoning. *Thinking & Reasoning*, *18*(3), 244–277.
- Henkel, L. A., & Mattson, M. E. (2011). Reading is believing: The truth effect and source credibility. *Consciousness and cognition*, *20*(4), 1705–1721.
- Hillygus, D. S. (2011). The evolution of election polling in the united states. *Public opinion quarterly*, *75*(5), 962–981.

- Kim, S., & Spelke, E. S. (2020). Learning from multiple informants: Children's response to epistemic bases for consensus judgments. *Journal of experimental child psychology*, *192*, 104759.
- Koch, T., & Zerback, T. (2013). Helpful or harmful? how frequent repetition affects perceived statement credibility. *Journal of Communication*, *63*(6), 993–1010.
- Kozyreva, A., Lewandowsky, S., & Hertwig, R. (2020). Citizens versus the internet: Confronting digital challenges with cognitive tools. *Psychological Science in the Public Interest*, *21*(3), 103–156.
- Kruskal, W. (1988). Miracles and statistics: The casual assumption of independence. *Journal of the American Statistical Association*, *83*(404), 929–940.
- Krynski, T. R., & Tenenbaum, J. B. (2007). The role of causality in judgment under uncertainty. *Journal of Experimental Psychology: General*, *136*(3), 430.
- Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge university press.
- Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide*. Psychology press.
- 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*.
- Maines, L. A. (1990). The effect of forecast redundancy on judgments of a consensus forecast's expected accuracy. *Journal of Accounting Research*, *28*, 29–47.
- Mercier, H., & Miton, H. (2019). Utilizing simple cues to informational dependency. *Evolution and Human Behavior*, *40*(3), 301–314.
- Morey, R. D., Rouder, J. N., Jamil, T., & Morey, M. R. D. (2015). Package 'bayesfactor'. URL <http://cran.r-project.org/web/packages/BayesFactor/BayesFactor.pdf> i (accessed 1006 15).

- Navon, D. (1978). The importance of being conservative: Some reflections on human bayesian behaviour. *British Journal of Mathematical and Statistical Psychology*, 31(1), 33–48.
- Nisbett, R. E., Krantz, D. H., Jepson, C., & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. *Psychological review*, 90(4), 339.
- Pilditch, T. D., Hahn, U., Fenton, N., & Lagnado, D. (2020). Dependencies in evidential reports: The case for informational advantages. *Cognition*, 204, 104343.
- Rabb, N., Fernbach, P. M., & Sloman, S. A. (2019). Individual representation in a community of knowledge. *Trends in cognitive sciences*, 23(10), 891–902.
- Ross, L., Greene, D., & House, P. (1977). The “false consensus effect”: An egocentric bias in social perception and attribution processes. *Journal of experimental social psychology*, 13(3), 279–301.
- Schum, D. A. (2001). *The evidential foundations of probabilistic reasoning*. Northwestern University Press.
- Schwarz, N., & Jalbert, M. (2020). When (fake) news feels true: Intuitions of truth and the acceptance and correction of misinformation. r. greifeneder, m. jaffé, ej newman, & n. schwarz. *The psychology of fake news: Accepting, sharing, and correcting misinformation*, 73–90.
- Schwarz, N., Sanna, L. J., Skurnik, I., & Yoon, C. (2007). Metacognitive experiences and the intricacies of setting people straight: Implications for debiasing and public information campaigns. *Advances in experimental social psychology*, 39, 127–161.
- Sedlmeier, P. (1999). *Improving statistical reasoning: Theoretical models and practical implications*. Psychology Press.
- Sulik, J., Bahrami, B., & Deroy, O. (2020). Social influence and informational independence. *Cognitive Science Conference Proceedings*, 19.

- Weaver, K., Garcia, S. M., Schwarz, N., & Miller, D. T. (2007). Inferring the popularity of an opinion from its familiarity: A repetitive voice can sound like a chorus. *Journal of personality and social psychology*, 92(5), 821.
- Weisberg, D. S., Keil, F. C., Goodstein, J., Rawson, E., & Gray, J. R. (2008). The seductive allure of neuroscience explanations. *Journal of cognitive neuroscience*, 20(3), 470–477.
- Whalen, A., Griffiths, T. L., & Buchsbaum, D. (2018). Sensitivity to shared information in social learning. *Cognitive science*, 42(1), 168–187.
- Winkler, R. L., & Murphy, A. H. (1973). Information aggregation in probabilistic prediction. *IEEE Transactions on Systems, Man, and Cybernetics*, (2), 154–160.
- Xie, B., Navarro, D. J., & Hayes, B. K. (2020). Adding types, but not tokens, affects property induction. *Cognitive Science*, 44(9), e12895.
- Yaniv, I., Choshen-Hillel, S., & Milyavsky, M. (2009). Spurious consensus and opinion revision: Why might people be more confident in their less accurate judgments? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(2), 558.
- 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.