Any consensus will do:
The failure to distinguish between 'true' and 'false' consensus

Sami R. Yousif (sami.yousif@yale.edu), Rosie Aboody (rosie.aboody@yale.edu), and Frank C. Keil (frank.keil@yale.edu)
Department of Psychology, Yale University
New Haven, CT, 06520 USA

Abstract

As we navigate our information-rich world, we frequently interpret and integrate testimony from external sources (friends, teachers, books, internet articles, etc.) – deciding which pieces of information to believe, and which to discard. One cue to a statement's trustworthiness is whether it comes from a *consensus* (i.e., when a majority of people agree). But what counts as consensus? When presented with a set of agreeing sources, do we evaluate the *quality* of consensus – for example, asking whether each source arrived at their conclusion by independent means? In a first experiment, we demonstrate that individuals are insensitive to the quality of a consensus, and are equally confident in conclusions drawn from a 'true' consensus (i.e., one derived from many primary sources) and those drawn from a 'false' consensus (i.e., one derived from many secondary sources but only a single primary source). In a second experiment, we find that this continues to be true even when the expertise of the secondary sources is minimized. Together, our experiments provide converging evidence that people do not properly discount (or discount at all) information from a 'false' consensus.

Keywords: consensus; conformity; social learning; reasoning

Introduction

How do we decide what information we can trust? In theory, at least, *consensus* is a fairly strong cue towards a claim's trustworthiness: if many sources all report the same story, it is reasonable to assume the story is more likely to be true. And indeed, a large literature has demonstrated that adults, children, and even dogs are highly sensitive to consensus (Asch, 1956; Corriveau & Harris, 2010; Kundey et al., 2012). But, is every consensus equally informative, or are there cases where a consensus appears superficially true, but is ultimately false?

Imagine, for example, that you hear a rumor from one of your colleagues that a brand-new coffee machine is to be installed in your department. A little while later, you hear the same rumor from another colleague, and then another. You have now heard the same information repeated several times, and there appears to be a clear consensus. While you could accept this information at face value, you could also probe deeper. For example, one might ask: where did your colleagues get *their* information from? If it turned out that all three had received their information from a single source (rather than from independent sources), would you be less likely to believe it?

While the above is clearly an example of a consensus, it does not seem to be a very *good* one. While there is one sense in which there clearly is some kind of consensus (in that multiple sources are conveying identical information), there is another sense in which this consensus is misleading: each source is merely repeating information they obtained from another source.

The above example highlights that consensus can refer to many different things, and that there is no single criterion that determines what counts as a consensus. But, insofar as these different kinds of consensus can be delineated, a question remains as to whether individuals distinguish between them – and if those distinctions have cognitive consequences.

The consequences of consensus

Classic work on conformity suggests that people may over-rely on apparent consensus (Asch, 1956), even when that consensus is obviously wrong. But, here, consensus refers only to agreement without any cause or explanation. More recently, sensitivity to consensus has been primarily studied in children. This work has shown that even young children are sensitive to consensus information, with children as young as 3 years old reliably aligning themselves with the majority when there is disagreement (Corriveau et al., 2009; Corriveau & Harris, 2010; Fusaro & Harris, 2008). Further work has investigated how and under what conditions children are susceptible to consensus (Burdett et al., 2016; Hu et al., 2013; Hu et al., 2015). Yet, little work has examined sensitivity to the quality of consensus (one recent paper being the exception; see Einav, 2018) – even though this question seems directly relevant to a wide range of psychological and sociological questions. This is especially important as we attempt to understand how it is that people come to believe (and defend) erroneous information. Indeed, consensus plays a critical part not only as we decide which rumors to take seriously, but also as we interpret academic articles, news sources, and virtually any kind of information for which there can be a consensus at all.

Other recent work has probed people's intuitions about the quality of consensus more directly. For example, recent work has shown that people will adopt sophisticated strategies rather than merely copying a majority. In particular, it was shown that participants were sensitive to whether testimony from many different sources was independent (Whalen et al., 2018). Thus, this work seems to predict that individuals would be sensitive to the difference between a true and false consensus.

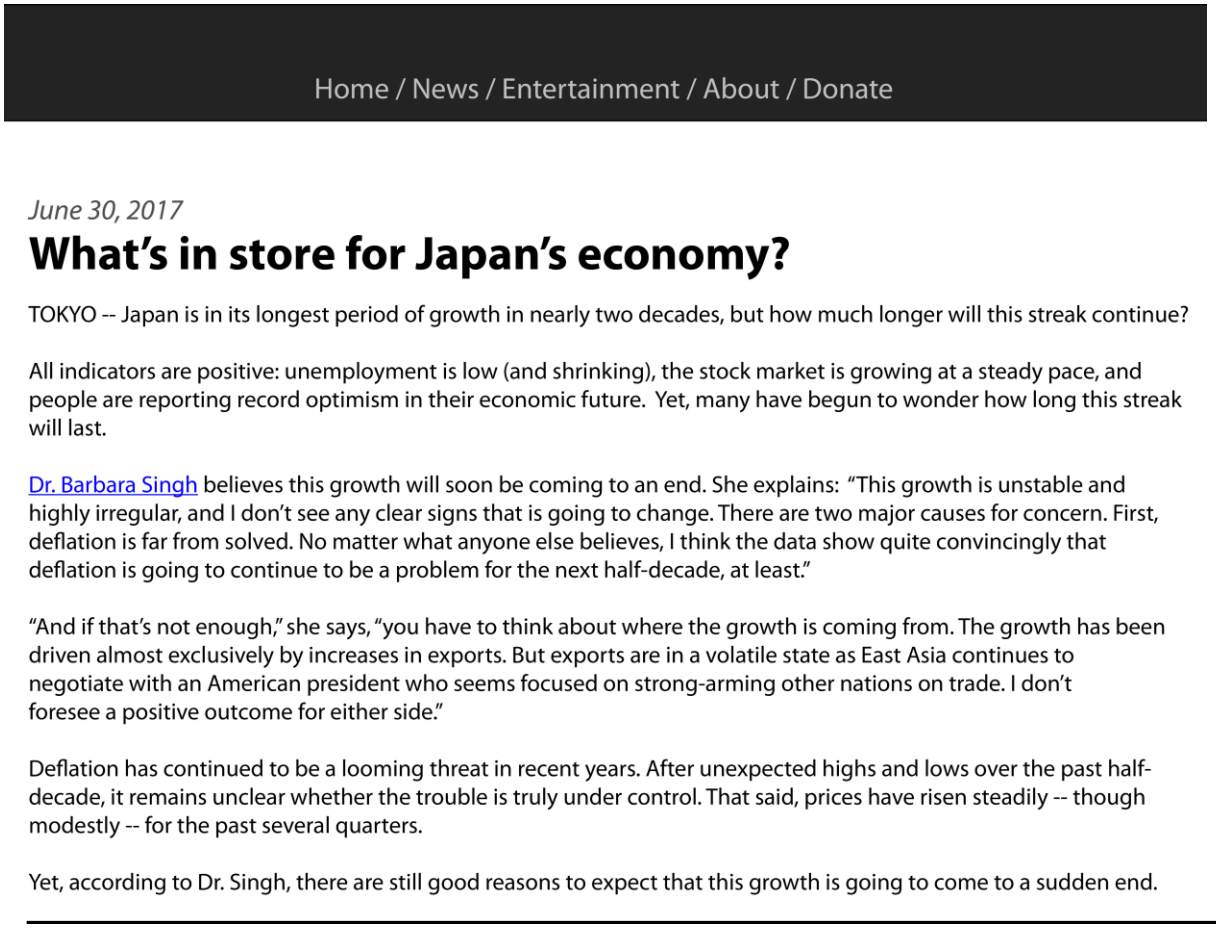

Figure 1. Sample article from Experiment 1.

Outside of the realm of cognitive science, one study has shed light on a timely example of 'false consensus' (and the consequences it may have), by investigating the sociology of climate change denial. The researchers wanted to understand why there is such a wide gap between the scientific consensus and public opinion. To do so, they investigated popular internet blogs devoted to discussions of climate change. Over 80% of the blogs studied relied on a single primary source – a single person who, despite having never conducted any research, is said to be an expert on (of all things) polar bears (Harvey et al., 2017). The work (or, opinion) of a single person is subsequently published over and over again in one source after another. There appears to be a clear consensus – and many people seem to take that consensus as probative. This sociological study seems to document an instance of *false* consensus. But perhaps laypeople reading such sources recognize false consensus when they see it, and properly discount the validity of information that is merely *repeated*. Alternatively, individuals may take repeated information *just as seriously* as information that comes from different primary sources (e.g. if, instead of citing primarily one source, the science-denying climate change blogs had relied on many *unique* sources).

Here, we test this question in two separate ways. First, we assess whether individuals are sensitive to the difference between 'true' and 'false' consensus in the first place, by having them read information from seemingly ordinary news articles (where, ostensibly, the secondary sources possess some level of expertise, or responsibility to report accurate information). Next, we assess whether individuals are sensitive to this difference when the secondary sources are explicitly non-expert (such that they might not have made any attempt to independently verify information). In both instances, we find that while individuals are sensitive to consensus (i.e., that they are more likely to believe this information as opposed to information for which there is no consensus), they fail to distinguish between true and false consensus (believing both equally as much).

Experiment 1: Assessing false consensus

We first assessed sensitivity to the quality of consensus in the most straightforward way possible: by simply showing participants artificial news articles, and assessing their confidence in the arguments presented (see Figure 1). To manipulate consensus, we varied a) the number of secondary sources that took a particular side in a debate and b) the number of primary sources those secondary sources cited.

Method

Participants 240 adult participants completed a survey online through Amazon Mechanical Turk (22 additional participants participated but were removed due to failing a simple attention check; see procedure). The sample size was chosen based on independent pilot data. All participants lived in the United States.

Stimuli Materials consisted of artificial news articles about the Japanese economy. These articles were written from scratch, although they were based on true information. Some articles took an affirmative stance (Japan's economy *will* continue to improve), and some took a negative stance (Japan's economy will *not* continue to improve). Participants were explicitly told that any markers of the articles' origins (as well as ads, etc.) had been removed to minimize distractions. Articles all cited their primary sources (the name of an expert), which was highlighted in blue and underlined as if it were a hyperlink. See Figure 1.

Procedure 80 participants each were randomly assigned to one of three conditions: a 'true consensus' condition, a 'no consensus' condition, and a 'false consensus' condition. In both the 'true' and 'false consensus' conditions, participants read one article that took a negative stance on the future of Japan's economy, and four articles that took a positive stance. The only difference between conditions was the number of primary sources these articles cited: In the 'true consensus' condition, each article cited a unique primary source. Critically, in the 'false consensus' condition, each of the positive articles cited *the same primary source*. And in the baseline 'no consensus' condition, participants read one article that took a negative stance and one article that took a positive stance. Each article cited a unique primary source. The articles were presented in a random order for each participant.

After reading the articles, participants were asked how much they agreed with the affirmative position (the one for which they read four articles in the 'true' and 'false' consensus conditions). They responded by clicking on a number line, indicating their confidence on a scale of 0-100. Participants were prompted to confirm their answer before submitting it. On a separate screen, participants were then asked two questions. First, they were asked which sources (from a list of 10) had been cited in the articles they read. They were free to select any number of sources that they wanted. Additionally, they were asked what nation the articles had been about, selecting their answer from a list of five possibilities. Participants who failed to answer the latter question correctly were excluded and replaced (true consensus: n = 10; false consensus: n = 4; no consensus: n = 8). No other information was collected from the participants.

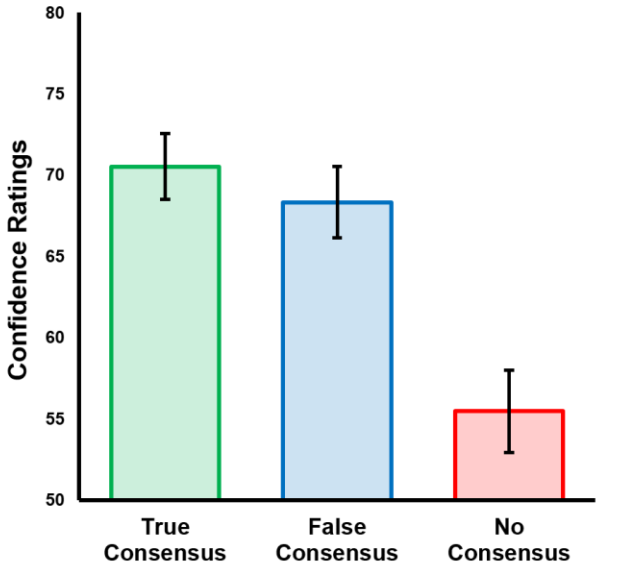

Figure 2. Results of Experiment 1: mean confidence ratings in the true, false, and no consensus conditions. Error bars represent +/- 1 SE.

Results and Discussion

First, it was critical to assess whether participants (as a whole) were tracking the individual sources that they read. We used a d-prime analysis to determine whether people correctly identified the sources they had seen, above what would be predicted by chance. In each condition, participants identified the correct primary sources at above-chance levels (*true consensus*: d'=1.16, $t(79)=8.63$, $p<.001$, $d=.97$; *no consensus*: d'=1.37, $t(79)=12.27$, $p<.001$, $d=1.37$; *false consensus*: d'=2.20, $t(79)=14.21$, $p<.001$, $d=1.60$)[1]. Thus, subsequent effects cannot be a result from failures to attend to (or remember) the sources in the first place.

The primary results of this experiment are reported in Figure 2. We first assessed whether consensus did in fact increase participants confidence in the information they read. Indeed, participants in the 'true consensus' condition were 15 points more confident than participants in the 'no consensus' condition (on a 100-point scale; $t(158)=4.71$, $p<.001$, Bonferroni corrected). But did participants discount consensus in the 'false consensus' condition? First, we assessed whether the 'true consensus' confidence ratings differed from those in the 'false consensus' condition. These were separated by only 2 points on a 100 point scale, and this difference was not significant ($t(158)=.69$, $p>.90$, Bonferroni corrected). Additionally, we tested whether confidence in the 'false consensus' condition was actually greater than in the 'no consensus' condition. Indeed, subjects were more confident in the 'false consensus' condition than in the 'no consensus' condition (amounting to a 13-point difference in confidence, or about a 23% increase; $t(158)=4.02$, $p<.001$, Bonferroni corrected). The

[1] For hit- or false-alarm-rates of 0 or 100%, we used values of 5% and 95% instead. Modifying these values to be as conservative as possible does not change the result.

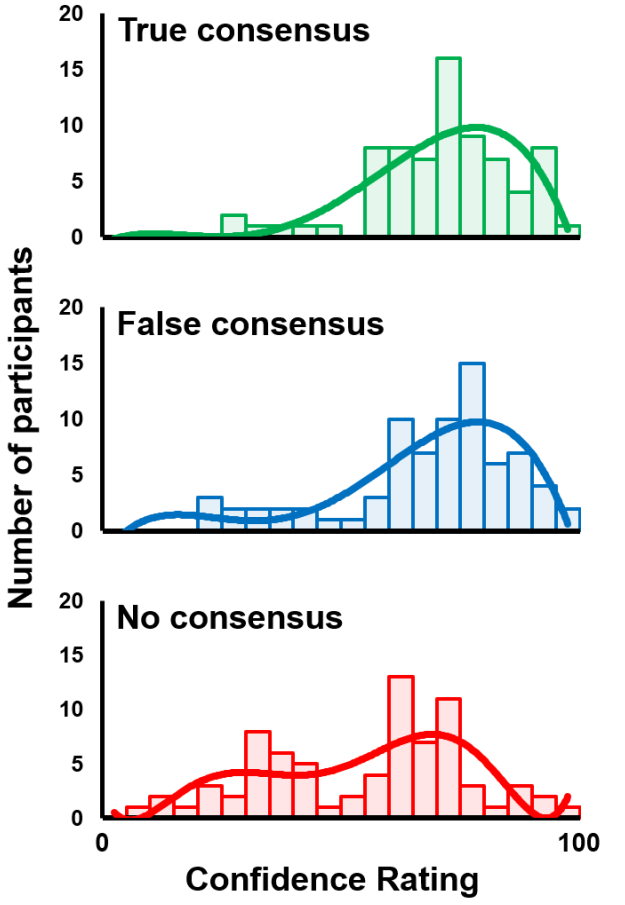

Figure 3. Results of Experiment 1: complete distributions of confidence ratings in the true, false, and no consensus conditions.

complete confidence rating distributions for each condition can be seen in Figure 3.

These results suggest that individuals are *not* sensitive to any difference between 'true' and 'false' consensus. While it is possible this effect could have been driven by subjects who failed to notice there was only one source in the false consensus condition, participants across all conditions were quite good at identifying which sources they had heard from. Furthermore, average confidence ratings in the true and false consensus conditions did not differ from one another. If the similarity of the true and false consensus conditions was being driven by only a few participants, one might expect this difference would be more pronounced. One way of testing this empirically is by removing participants whose d-primes were not higher than zero (i.e., participants who were particularly poor at identifying the sources they had heard from). Doing so actually makes our effect *even more pronounced*: the difference between 'false consensus' and 'no consensus' increases to 16 points (amounting to more than a 30% increase in confidence).

Alternatively, there may be a rational explanation for these results. When reading news articles, it is possible that individuals assume news sources possess some amount of journalistic integrity (whether or not this happens to be true). If twenty different sources are reporting on rumors from unnamed sources in the White House, perhaps it is fair to assume that each source verified those claims to some degree – even if they all cited a single source. Furthermore, there is sense in which it may not only be reasonable for secondary sources to cite a single primary source, but that it may simply be the ethical thing to do (i.e., giving credit where credit is due).

Might individuals discount false consensus if sources are explicitly non-expert? Experiment 2 manipulated expertise to address this concern.

Experiment 2: Minimizing expertise

We next assessed sensitivity to consensus in a case where participants were unlikely to make assumptions about the knowledge and expertise of the secondary sources. To do so, we had participants read artificial student essays rather than news articles. The purpose of this manipulation was to make it clear to participants that the secondary sources had not – and indeed, could not – have independently verified the claims they were asserting.

Method

All elements of the experimental design were identical to those of Experiment 1, except as stated below. 240 new participants the survey online through Amazon Mechanical Turk (and 59 additional participants participated but were removed for failing a simple attention check; true consensus: n = 22; false consensus: n = 16; no consensus: n = 21).

Instead of reading news articles about the state of the Japanese economy, participants read artificial student essays about an impending tax proposal in Sweden. Prior to reading the articles, participants were told explicitly (i.e., in bold letters, in the center of the screen) that the students had been specifically instructed to cite their sources in order to make their respective arguments. Unlike the sources in the previous experiment, the primary sources in this experiment were the names of (real) economic foundations. Again, the articles could take either a negative stance (the tax policy should *not* be approved) or an affirmative stance (the tax policy *should* be approved).

Results and Discussion

Again, it was critical to assess whether individuals (as a whole) were tracking the individual sources cited in the essays. And, again, this turned out to be true. We used a d-prime analysis to determine whether people correctly identified the sources they had seen, above what would be predicted by chance. In each condition, it was true that subjects identified the correct primary sources at above-chance levels (*true consensus*: d'=1.66, $t(79)=12.87$, $p<.001$,

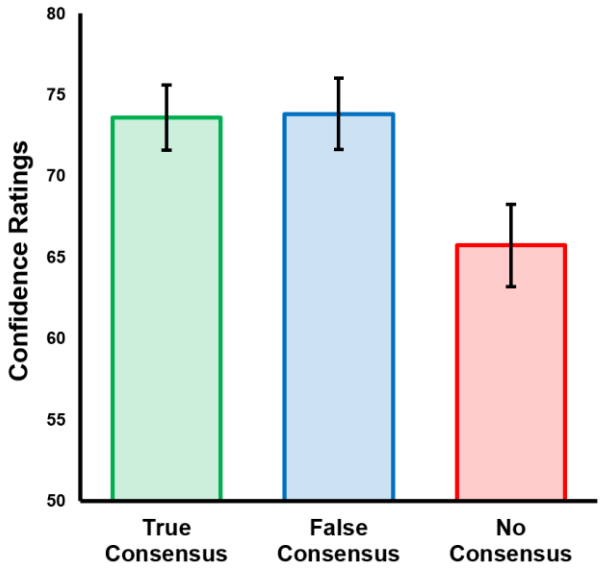

Figure 4. Results of Experiment 2: mean confidence ratings in the true, false, and no consensus conditions. Error bars represent +/- 1 SE.

d=1.44; *no consensus*: d'=1.37, *t*(72)=5.20[2], *p*<.001, *d*=.61; *false consensus*: d'=1.16, *t*(79)=7.13, *p*<.001, *d*=.80). Thus, we can be confident that any subsequent effects did not result from a failure to attend to (or remember) the sources in the first place.

The primary results of this experiment are reported in Figure 4. Analyses confirm what is evident from the figure. Consensus did in fact increase participants' confidence in the information they read. Indeed, participants in the 'true consensus' condition were 8 points more confident than participants in the 'no consensus' condition (on a 100-point scale; *t*(158)=2.42, *p*<.05, Bonferroni corrected). Again, we also assessed whether the 'true consensus' confidence ratings differed from those in the 'false consensus' condition. The two were separated by only two-tenths of a point on a 100 point scale, and this difference was not significant (*t*(158)=.07, *p*>.90, Bonferroni corrected). Participants were also more confident in the 'false consensus' condition than in the 'no consensus' condition (amounting to a 8-point difference in confidence, or about a 13% increase; *t*(158)=4.02, *p*<.001, Bonferroni corrected).

Once again, we tested whether these effects could be explained by some subset of the participants who did not attend to source information. After removing participants with d-primes less than or equal to zero, the effects were once again even stronger: the difference between the 'true consensus' and 'no consensus' condition was still 8 points, and the difference between 'false consensus' and 'no consensus' increased to 9 points (amounting to about a 15% increase in confidence).

[2] Due to missing data caused by subjects refreshing the page in error, seven subjects were excluded from this analysis.

In sum, these results provide converging evidence that individuals do not properly discount information that is merely repeated. In fact, participants' trust in repeated information is exactly the same as their trust in independently sourced information.

General Discussion

In a first experiment, we showed that individuals fail to differentiate true and false consensus when exposed to ordinary news articles. In a second experiment, we showed that this effect persists even when minimizing the expertise of the secondary sources. Together, these results suggest that individuals are sensitive to consensus only in a superficial way – and that genuine consensus confers no additional increase in confidence.

How can these results be explained? Individuals seem to be aware that they are only (truly) hearing from one source, and yet they are still quite confident in the information they read. Below, we highlight five considerations for future work.

Other considerations and future directions

First: our work is not meant to answer any philosophical or sociological question about whether we *ought* to discount false consensus. We can imagine arguments on both sides. Rather, we simply want to understand *how* people interpret a consensus, and what counts as a consensus in the first place – regardless of whether this is rational, or sensible.

Second, although individuals as a whole are tracking the sources at above-chance levels, it is challenging to determine on a participant basis who exactly paid attention to the names of the sources, and whether (or how) this may have biased our results. For example, the higher confidence ratings in the false consensus condition could have been driven by a subset of participants who attended to the number of articles, but not the number of primary sources. If true, this might mean that the increased confidence in the false consensus condition (compared to the no consensus condition) was driven solely by the subset of participants who had an incorrect sense of the number of sources they had heard from.

There are two reasons to doubt this explanation. First, confidence in the false consensus condition was not only *higher* than confidence in the no consensus condition: it was also equal to confidence in the true consensus condition. Therefore, it is unlikely that a small number of participants drove this effect. But more importantly, when we excluded participants who had a zero-or-lower d-prime score (indicating they had not attended to the sources), our effects become even more pronounced. Thus, our results do not seem to be driven by noisy responses, but diminished by them. Moreover, whether we used this exclusion criteria or other comparable ones, our results were always qualitatively the same.

Third, it remains unclear *how* participants are interpreting the secondary sources. Even in our second experiment, which was designed to minimize assumptions made about

the secondary sources (students), it is quite possible that individuals assume these sources acted as some kind of filter: perhaps they all attended to (and wrote about) the same primary source because that information had been most persuasive to them to begin with. Future work may address this by studying the phenomenon of 'false consensus' in instances where the secondary sources *could not possibly* have outside knowledge.

Fourth, we did not explore other conditions under which individuals might properly discount false consensus. We very purposefully directed attention to the sources in each experiment, and participants in different conditions were able to identify which source(s) they had heard from. However, it is possible that an even more heavy-handed approach would have helped participants to recognize and discount instances of false consensus (as in Whalen et al., 2018). For example, what if participants were prompted to recall what sources they would heard from *before* their confidence had been assessed? Would this cause them to reason differently about the trustworthiness of the false consensus?

Finally, it remains unclear how robust these effects would be across contexts. Our examples (one about the Japanese economy, and the other about Swedish tax policy) purposefully straddled the fence between objectivity and subjectivity. That is, arguments of taxation and economics are, in principle, grounded in empirical data, yet there is room for interpretation. Other work has shown that people interpret consensus differently depending on the level of perceived subjectivity/objectivity of information (Yousif & Keil; under review); might such factors also play a role as people interpret the quality of a consensus?

Conclusion

Overall, our results suggest that individuals do not discount information that is merely repeated. Instead, individuals seem just as confident in information derived from a false consensus as they are in information derived from a true consensus. It is not hard to imagine, in light of these results, why fake news articles spread across social media may have such powerful effects on political outcomes. Yet, these findings also raise important questions for the future. How can we combat this over-reliance on consensus? Should news sources differentiate between claims that are being *repeated* and claims that have been *verified* – and would this even make a difference?

In sum, while consensus is a powerful sociological signal, it may not always be the best empirical signal. Yet, it seems that people may not notice the difference – and this cognitive neglect has serious implications for how people perceive and interpret the over-abundance of information they are exposed to on a daily basis.

Acknowledgments

For helpful comments and conversation, we thank Clara Colombatto, Brynn Sherman, and all the members of the Yale Cognition and Development Lab.

References

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

Burdett, E. R., Lucas, A. J., Buchsbaum, D., McGuigan, N., Wood, L. A., & Whiten, A. (2016). Do children copy an expert or a majority? Examining selective learning in instrumental and normative contexts. *PloS One*, *11*(10).

Corriveau, K. H., Fusaro, M., & Harris, P. L. (2009). Going with the flow: Preschoolers prefer nondissenters as informants. *Psychological Science*, *20*, 372-377.

Corriveau, K. H., & Harris, P. L. (2010). Preschoolers (sometimes) defer to the majority in making simple perceptual judgments. *Developmental Psychology*, *46*, 437.

Einav, S. (2018). Thinking for themselves? The effect of informant independence on children's endorsement of testimony from a consensus. *Social Development*. *27*, 73-86.

Fusaro, M., & Harris, P. L. (2008). Children assess informant reliability using bystanders' non-verbal cues. *Developmental Science*, *11*, 771-777.

Harvey, J. A., van den Berg, D., Ellers, J., Kampen, R., Crowther, T. W., Roessingh, P., ... & Stirling, I. (in press). Internet blogs, polar bears, and climate-change denial by proxy. *BioScience*.

Hu, J. C., Buchsbaum, D., Griffiths T. L., & Xu F. (2013). When does the majority rule? Preschoolers trust in majority informants varies by domain. Proceedings of the 35th annual conference Cognitive Science Society (pp. 2584-2589), Austin, TX: Cognitive Science Society.

Hu, J. C., Whalen, A., Buchsbaum, D., Griffiths, T. L., & Xu, F. (2015). Can children balance the size of a majority with the quality of their information? Proceedings of the 37th annual conference Cognitive Science Society (pp. 956-961), Austin, TX: Cognitive Science Society.

Kundey, S. M., German, R., De Los Reyes, A., Monnier, B., Swift, P., Delise, J., & Tomlin, M. (2012). Domestic dogs' (Canis familiaris) choices in reference to agreement among human informants on location of food. *Animal Cognition*, *15*, 991-997.

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

Yousif, S. R. & Keil, F. C. (under review). When in doubt: Using confidence and consensus as 'summary statistics' of collective knowledge.