

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

When in doubt: Using confidence and consensus as ‘summary statistics’ of collective knowledge

Permalink

<https://escholarship.org/uc/item/2n76559q>

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 40(0)

Authors

Yousif, Sami R

Keil, Frank C

Publication Date

2018

When in doubt: Using confidence and consensus as ‘summary statistics’ of collective knowledge

Sami R. Yousif (sami.yousif@yale.edu) and Frank C. Keil (frank.keil@yale.edu)

Department of Psychology, Yale University

New Haven, CT, 06520 USA

Abstract

People do not think in isolation. Whether purchasing a new product on Amazon, deciding what movie to watch, or evaluating scientific evidence, we often rely on aggregated sources of information (e.g., product ratings or reviews) to make decisions. Indeed, the internet has given rise to unprecedented levels of aggregated information, to the extent that it is difficult to imagine anything for which we might *not* be able to find summary information. In other words, what we know (or think we know) is constrained not just by our own knowledge, but by the knowledge of our *community* (Sloman & Rabb, 2016). Yet this raises a question: what happens when a community of knowledge is not in agreement? Here, we assess this question by pitting cases of *high confidence* against cases of *high consensus*. Results from two experiments show that 1) individuals are sensitive to both confidence and consensus; 2) individuals utilize such information in a predictable but context-dependent manner; and 3) perceptions of confidence and consensus influence judgments and decisions in a substantial way, even when individuals are not aware of the contrast between them. Taken together, the findings suggest that individuals are highly sensitive to variability in aggregated information – rather than merely an average – and that these ‘summary statistics’ of aggregated information have a substantial, reliable impact on decision-making.

Keywords: decision-making; reasoning; summary statistics; confidence; consensus

Introduction

Imagine that you are interested in purchasing a new book. Rather than selecting the book with the most enticing cover, you might instead consult some online reviews. In particular, you might rely on certain summary statistics in the form of Likert-esque ratings. If the reviews are consistently positive, then you would have no trouble making a decision. But what if there is inconsistency? Assume you have narrowed down your search to two books: one of them has consistently mediocre reviews, while the other has a mix of highly positive and highly negative reviews. Which book would you choose? Under certain conditions, at least, you might select the book with higher variability in its ratings (Sun, 2012). This bias may occur because people are more drawn to a positive experience than they are averse to a negative one (e.g., see Dodds et al., 2015). After all, nothing is stopping you from setting the book down if you’re not enjoying it. Or perhaps people are likely to discount highly negative ratings in the presence of other positive ratings. If many people enjoy a product,

perhaps those who do not are just picky or have different tastes.

But now consider a different example: you are reading a scientific article about a particular topic for which there are two competing theories. All scientists agree that one theory (Theory A) is about 50% likely to be true. Alternatively, 50% of scientists think that the other theory (Theory B) must be correct, and the other 50% of scientists think that it (Theory B) could not possibly be correct. If you had to, would you opt for high confidence (the latter) or high consensus (the former)? More importantly: would you make the same kind of decision in both the objective scenario (choosing between two scientific theories) and the subjective scenario (choosing between two books)?

Confidence versus consensus

The present work explores confidence and consensus as two ‘summary statistics’ of aggregated information. Though never contrasted in this way (to our knowledge), confidence and consensus have been studied in other areas of research. For example, prior work has identified how and why individuals are overconfident in the first place (Dunning, 2012; Dunning et al., 2003; Mills & Keil, 2005), and how assessments of confidence can influence courtroom decisions (Sporer, et al., 1995; Wells et al., 1979) and beyond (e.g. Estes & Hosseini, 1988; Park & Park, 2013; Snizek, 1992; West & Broniarczyk, 1998). Even children make sophisticated inferences about the expertise of third-parties based on their level of confidence (Kominsky, Langthorne, & Keil, 2016), and understand that a lack of consensus means something different in matters of taste versus matters of fact (Wainryb et al., 2004).

In addition, classic studies of conformity reveal that individuals consistently side with the majority when there is consensus on a particular issue, even when that majority appears to disagree with an obvious, objective truth (Asch, 1956). More recent work has shown that even young children are sensitive to consensus information, with 3-year-olds reliably aligning themselves with the majority when there is disagreement (Corriveau et al., 2009; Fusaro & Harris, 2008; for a more nuanced view, see: Einav, 2018). Additional work has shown that children do not only rely on consensus, but that they may occasionally prefer, for example, an expert opinion over a majority (Burdett et al., 2016). This work demonstrates that consensus may not always be the best cue to the quality of information. Thus, the present work aims to understand the trade-off between

Table 1: Example Objective Scenarios

Category	Vignette
Science	You have been selected to help choose between two new technologies which could be integral in helping to send spacecraft to Mars. Two teams of scientists have been independently working on a new means of jet propulsion. However, the scientific community is uncertain: 100% of scientists believe that Option A has only a 50% probability of being successful; whereas 50% of scientists think that Option B has a 100% chance of being successful and the other 50% of scientists think that Option B has a 0% chance of being successful.
Health	You have been selected to help choose between investing in one of two new antibiotics which could help to eradicate many ailments across the globe. Two teams of biochemists have been independently working on two separate antibiotics. However, the expert community is uncertain: 100% of experts believe that Option A has only a 50% probability of being successful; whereas 50% of experts think that Option B has a 100% chance of being successful and the other 50% of experts think that Option B has a 0% chance of being successful.

Note: These are example scenarios for Experiment 1A only. Control experiments used variants of these vignettes.

consensus and another dimension we may use to assess collective information: confidence.

For our purposes, confidence refers not to an internal state of certainty or uncertainty, but rather to a projected certainty or uncertainty regardless of any underlying mental state. This distinction is critical in the present work, because we are studying how people reason about confidence and consensus as a third party interpreting summary information. In the real world, when interpreting information from a distance, we rarely have access to mental state information. This is akin to studies of consumer behavior (e.g., Park & Park, 2013; West & Broniarczyk, 1998), where consensus has been contrasted with polarity (which roughly maps on to our idea of confidence).

Current Study

Adults (Spiller & Belogolova, 2016) and children (Wainryb et al., 2004) differentiate between matters of taste and matters of fact. Might this mean we also interpret collective information in these two domains differently? Two sets of experiments evaluated how individuals use confidence and consensus to guide decisions in different contexts – namely, those that involved matters of fact, and those that involved matters of taste. In the first experiment, naïve participants made a dichotomous choice in one of four, objective scenarios (that is, scenarios for which there is ostensibly some underlying truth; see Table 1). In each case, participants were told to select one of two possible options: a case of a high confidence or a case of high consensus. In a follow-up experiment, participants made a similar evaluation for subjective scenarios (that is, scenarios for which there is no underlying truth; see Table 2). Here, too, participants selected between options of high confidence and high consensus. Finally, in a second set of studies, we

exposed participants to *either* a high confidence option *or* a high consensus option, and asked them to rate the extent to which they thought each option would be both factual and enjoyable.

Here, we find evidence that people’s use of confidence and consensus is highly context-dependent – such that individuals reliably prefer confidence in matters of taste and consensus in matters of fact. Furthermore, our studies show not only that individuals are sensitive to the distinction between confidence and consensus, but that it has a substantial impact on subsequent decision-making.

Experiment 1A: Objective scenarios

Do individuals prefer confidence or consensus when forced to choose between the two? We first assessed this question in the most direct way possible: by exposing participants to objective, scientific scenarios and asking them to evaluate various debates.

Method

388 adult participants completed a survey online through Amazon Mechanical Turk. 12 additional participants were removed for failing an attention check. The sample size was chosen based on independent pilot data. All participants lived in the United States.

Participants were exposed to one of four scenarios and asked to make a dichotomous choice regarding those scenarios (randomly determined; see Table 1). In each scenario, participants were asked to evaluate a debate. In particular, they were told that experts were choosing between two possible options: an option with high consensus (but low confidence) and an option with high confidence (but low consensus). All four scenarios involved objective scientific scenarios for which there was

Table 2: Example Subjective Scenarios

Category	Vignette
Book	Imagine that you are interested in purchasing a book, and you have to decide between one of two possible options. Both of these books cost the same amount, are written by the same author, and are of the same genre. 100% of people have said that Option A is average. If asked to rate the book, people consistently rate it as being 5 out of 10. On the other hand, 50% of people think that Option B is exceptional, rating it to be a 10 out of 10. However, the other 50% think that Option B is terrible, rating it to be a 0 out of 10.
Museum	Imagine that you are interested in visiting a museum, and you have to decide between one of two possible options. Both of these museums are free to visit, have a similar number of exhibits, and a diverse range of content. 100% of people have said that Option A is average. If asked to rate the museum, people consistently rate it as being 5 out of 10. On the other hand, 50% of people think that Option B is exceptional, rating it to be a 10 out of 10. However, the other 50% think that Option B is terrible, rating it to be a 0 out of 10.

Note: These are the example scenarios for Experiment 1B only. Control experiments used variants of these vignettes. The most critical variant of this experiment used exclusively probabilities rather than ratings, as in those in Table 1.

(ostensibly) a clear, measurable outcome. Diverse categories ensured that any observed effects would be general in scope. After reading the scenario, participants were told that ‘A decision needs to be made quickly, and you must choose one of the options. Which would you choose?’. They then made their selection.

Participants were asked two questions about their general level of expertise with each topic. They were asked 1) how much they know about the topic and 2) how easy it would be for them to learn about the topic if they had to. Because level of expertise was unrelated to our findings (in this experiment and subsequent experiments), these questions are not mentioned in the results. Participants also provided basic demographic information.

Results and Discussion

Across all four scenarios, participants opted for high consensus significantly more often than high confidence (256 out of 388 selections; binomial test, $p < .001$). Critically, this same pattern of results was independently true for each of the four scenarios (science: $p < .001$; humanitarian: $p = .025$; health: $p < .001$; business: $p = .003$; see Figure 1A).

These results suggest that participants preferred consensus over confidence. One noteworthy aspect of the vignettes tested here is the presence of risk, some (health) even concerning the loss of human life. Does risk alone explain why participants opt for consensus in objective scenarios? A series of control experiments which explicitly manipulated and minimized risk ruled out this possibility (though not included here, the results of those experiments are qualitatively identical to those reported here).

But do individuals *always* prefer consensus over confidence? Experiment 1B explores the possibility that

individuals’ tendency to rely on confidence or consensus may be context-dependent.

Experiment 1B: Subjective scenarios

Consider the two different scenarios posed in the introduction. In one case, you were asked to choose between two scientific theories (much like our vignettes from Experiment 1A). In another case, you had been asked to imagine purchasing a new book. Would you also opt for consensus in this case? Experiment 1B tested the hypothesis that when reasoning about *subjective* decisions (like choosing which book to read), individuals might instead prefer confidence over consensus.

Method

356 adult participants completed the survey online through Amazon Mechanical Turk. 44 additional participants were removed for failing an attention check. The sample size was chosen to be identical to that of the prior experiment (before exclusions). All participants lived in the United States.

All elements of the experimental design were identical to those of Experiment 1A, except that participants were now exposed to *subjective* scenarios (for which there is ostensibly no underlying, verifiable truth; see Table 2). The vignettes covered a diverse range of scenarios to ensure generality of results. Unlike the previous experiments, participants were given ratings of the choices rather than probabilities, to make the scenarios more familiar to the participants (who, like you, may frequently make choices based on information in this sort of rating format). However, we ran this exact experiment using both ratings and probabilities (just like those used in Experiment 1A). The results were qualitatively identical (and, for that matter, quantitatively identical), no matter which scale we used.

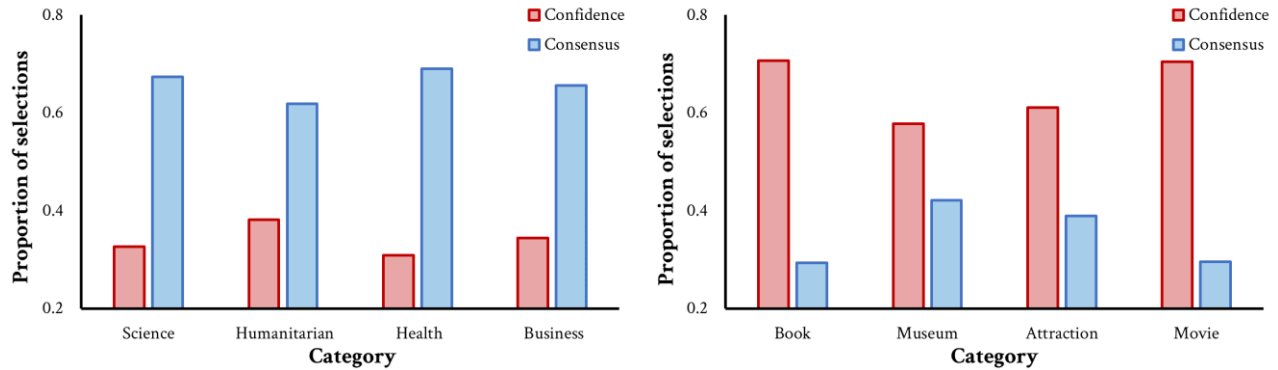


Figure 1. The proportion of ‘confidence’ (depicted in red) and ‘consensus’ (depicted in blue) choices for each of the four categories tested in (a) Experiment 1A (objective scenarios) and (b) Experiment 1B (subjective scenarios).

Results and Discussion

Across all four scenarios, participants opted for high confidence significantly more often than high consensus (266 out of 356 selections; binomial test, $p < .001$). Note: this is the *opposite* pattern from what we observed in Experiments 1A and 1B. Critically, this same pattern of results was independently true for each of the four scenarios (book: $p < .001$; museum: $p < .001$; attraction: $p = .007$; movie: $p < .001$; see Figure 1B).

Whereas individuals consistently preferred consensus over confidence in objective scenarios, the opposite is true for subjective scenarios. Indeed, this effect was present in each of the four scenarios tested.

These scenarios differed from the previous scenarios in two ways. First, as noted previously, we slightly modified the response scheme. Rather than using probabilities to differentiate between confidence and consensus, we used ratings. Additionally, these scenarios carried virtually no risk. Whereas a decision about a drug that could save human lives carries obvious risk, selecting which book to read carries virtually none. Just as with the prior experiment, we ran a series of supplemental experiments to address each of these possibilities. Neither of these factors altered the tendency to prefer confidence in matters of taste: the results were qualitatively the same as those reported here.

Experiment 2: Between-subjects

The two prior experiments pit confidence and consensus against one another by forcing participants to make a dichotomous choice. While the dissociation between confidence and consensus is striking, it is somewhat artificial and may not capture natural intuitions: how often do we choose between such clearly contrasted options in the real world? To address this concern, we exposed individual subjects to *either* a high confidence *or* a high consensus option (in the form of product ratings for a textbook). We then asked them to rate how enjoyable and how factual they thought the book would be.

Method

174 adult participants completed the survey online through Amazon Mechanical Turk. 26 additional participants were removed for failing an attention check. The sample size was chosen based on the prior studies (each prior experiment had 100 participants/condition, but here there are only two conditions). All participants lived in the United States.

Participants were told that they were going to be asked some questions about a textbook. They could be in one of two conditions. Either they were told that experts exhibited high confidence but low consensus (i.e. “half of them think the book is exceptional... however, the other half think that it is terrible”), or vice versa (i.e. “virtually all of them rated the book to be average”). They then evaluated the book’s potential enjoyability and factualness on a Likert scale.

Results and Discussion

Results from this experiment are depicted in Figure 2. As is evident from the figure, participants in the high confidence condition rated the potential enjoyability of the textbook to be higher than did participants in the high consensus group, $t(172) = 4.84$, $p < .001$, $d = .73$. Conversely, participants in the high consensus condition rated the factualness of the textbook to be higher than did participants in the high confidence group, $t(172) = 2.10$, $p = .037$, $d = .32$.

Overall, these results replicate the results of the previous experiments, but in a very different context. Here, subjects were exposed to confidence and consensus options independently, meaning that they were unaware of the alternative. Nevertheless, participants still rated the high confidence option as more enjoyable and the high consensus option as more factual – mapping on to the object/subjective distinction we explicitly manipulated between Experiments 1A and 1B. This result converges with the prior results because a) it demonstrates a dissociation between confidence and consensus and b) it shows that this dissociation differentially impacts objective and subjective measures.

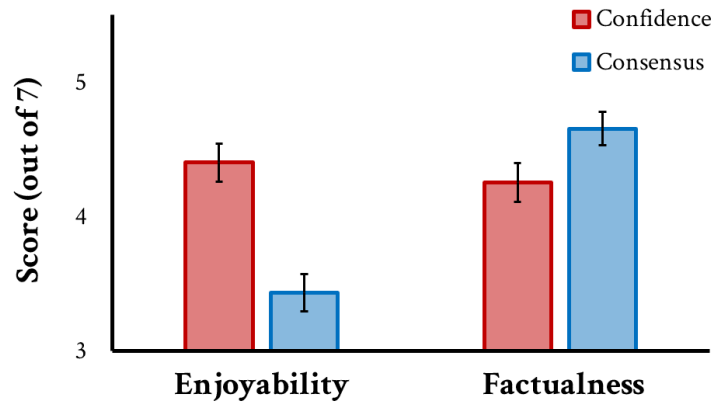


Figure 2. The mean ‘enjoyability’ and ‘factualness’ scores for each condition tested in Experiment 2. Error bars represents +/- 1 SE.

The between-subjects design of this experiment rules out the possibility that the prior results are due to some subtle, artificial feature of the contrast we created, while also ruling out the possibility that participants differentiate between confidence and consensus only when they are aware of the contrast. Indeed, the level of polarity in the information – even in such a minimalistic scenario – had a substantial impact on subsequent decisions. These patterns have clear implications for factors influencing an individual’s willingness to purchase a product, to decide which restaurant to visit, or to adopt a perspective on a contentious topic.

General Discussion

The present work has shown that individuals use confidence and consensus information in a predictable but context-dependent manner. Specifically, individuals prefer consensus over confidence in objective scenarios (Experiment 1A), but confidence over consensus in subjective scenarios (Experiment 1B). A separate set of experiments revealed how the confidence/consensus distinction influences behavior in a substantial way *even when individuals are unaware of the alternative* (Experiment 2).

Confidence and consensus in the real world

Consider again how often you consult reviews on Amazon or Yelp, or read a paper that summarizes two opposing perspectives, or offer your opinion on a highly politicized matter. In all of these cases, both explicitly and implicitly, you are making inferences about the patterns of confidence and consensus across your information sources. If a scientific theory (in an unfamiliar domain) is highly contentious, perhaps you would be reluctant to believe strongly in either perspective. However, if a book has highly variable reviews, perhaps you would still be willing to give the book a chance. What explains this dissociation?

Though we did not formally code our participants’ explanations, we did ask them why they made the decisions

that they did. In the objective scenarios, people seemed confused by how the scientific community could be so divided over something that presumably relied on some objective truth. So, they opted for consensus. In the subjective scenarios, by contrast, individuals seemed to discount the extremely negative information in the confidence cases. They would say things like ‘Someone enjoyed it, so those who don’t must just be especially picky’. Perhaps when we feel qualified to have our own opinions, it is easier for us to discount extreme opinions. When we are less sure, however – because we know that there must be an objectively correct answer – we like to rely on the consensus of the group.

In part, these findings help to explain – and possibly predict – behavior in the real world. If nothing else, this work demonstrates that individuals making decisions are sensitive to far more than the *average* opinion. Indeed, individuals are highly sensitive to the level of variability in opinions – and this sensitivity might well influence an individual’s decision to make a purchase, choose a destination, or form an opinion for themselves.

These findings may also inform how we ought to approach contentious topics. Consider the particularly (albeit mysteriously) divisive issue of climate change: that faced with overwhelming disagreement, individuals become only more certain of their own opinions. In other words, climate change seems like a clear example of the confidence effect documented here – despite the fact that there is some underlying truth. Perhaps this tells us something about how some people perceive the debate in the first place: not as a matter of science, but as a matter of personal judgment. Indeed, the issue is sometimes presented as one with two distinct positions (as opposed to one based on some objective truth, known or not; see, e.g., Harvey et al., 2017).

Moreover, these results reveal how context influences the processing of collective information. Though previous work has examined highlighted the distinction between matters of fact and matters of taste (e.g., Spiller & Belogolova), little is known about how people make sense of the information they receive in each of these domains. The present work

reveals a clear dissociation between confidence/consensus and matters of taste/matters of fact. Future work may benefit from further unpacking the exact cause of this dissociation.

In sum, confidence and consensus are two critical summary statistics used to evaluate aggregated information. Individuals reliably prefer confidence in matters of taste and consensus in matters of fact. But beyond simply choosing between confidence and consensus, individuals spontaneously use confidence levels as a cue: the distinction between confidence and consensus matters, even when no explicit contrast is made between them. Combined, these heuristics help explain how we make decisions while immersed in communities of knowledge that do not always know the answer, or may sometimes disagree.

Acknowledgments

For helpful comments and conversation, we thank Rosie Aboody, Matt Fisher, 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), e0164698.
- Conover, M., Ratkiewicz, J., Francisco, M. R., Gonçalves, B., Menczer, F., & Flammini, A. (2011). Political polarization on twitter. *ICWSM*, 133, 89-96.
- Corriveau, K. H., Fusaro, M., & Harris, P. L. (2009). Going with the flow: Preschoolers prefer nondissenters as informants. *Psychological Science*, 20, 372-377.
- Dodds, P. S., Clark, E. M., Desu, S., Frank, M. R., Reagan, A. J., Williams, J. R., ... & Megerdumian, K. (2015). Human language reveals a universal positivity bias. *Proceedings of the National Academy of Sciences*, 112, 2389-2394.
- Dunning, D. (2012). Confidence considered: Assessing the quality of decisions and performance. In P. Briñol & K. G. DeMarree (Eds.), *Social Metacognition* (63-80). New York: Taylor & Francis.
- Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science*, 12, 83-87.
- 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.
- Estes, R., & Hosseini, J. (1988). The gender gap on Wall Street: An empirical analysis of confidence in investment decision making. *The Journal of Psychology*, 122, 577-590.
- Harvey, J. A., van den Berg, D., Ellers, J., Kampen, R., Crowther, T. W., Roessingh, P., ... & Stirling, I. (2017). Internet blogs, polar bears, and climate-change denial by proxy. *BioScience*.
- Fusaro, M., & Harris, P. L. (2008). Children assess informant reliability using bystanders' non-verbal cues. *Developmental Science*, 11, 771-777.
- Kominsky, J. F., Langthorne, P., & Keil, F. C. (2016). The better part of not knowing: Virtuous ignorance. *Developmental Psychology*, 52, 31-45.
- Mills, C. M., & Keil, F. C. (2005). The development of cynicism. *Psychological Science*, 16, 385-390.
- Park, S. B., & Park, D. H. (2013). The effect of low-versus high-variance in product reviews on product evaluation. *Psychology & Marketing*, 30, 543-554.
- Prior, M. (2013). Media and political polarization. *Annual Review of Political Science*, 16, 101-127.
- Sloman, S. A., & Rabb, N. (2016). Your understanding is my understanding: evidence for a community of knowledge. *Psychological science*, 27, 1451-1460.
- Snizek, J. A. (1992). Groups under uncertainty: An examination of confidence in group decision making. *Organizational Behavior and Human Decision Processes*, 52, 124-155.
- Spiller, S. A., & Belogolova, L. (2016). On consumer beliefs about quality and taste. *Journal of Consumer Research*, 43, 970-991.
- Sporer, S. L., Penrod, S., Read, D., & Cutler, B. (1995). Choosing, confidence, and accuracy: A meta-analysis of the confidence-accuracy relation in eyewitness identification studies. *Psychological Bulletin*, 118, 315-327.
- Sun, M. (2012). How does the variance of product ratings matter? *Management Science*, 58, 696-707.
- Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *The quarterly journal of economics*, 106, 1039-1061.
- Wainryb, C., Shaw, L. A., Langley, M., Cottam, K., & Lewis, R. (2004). Children's thinking about diversity of belief in the early school years: Judgments of relativism, tolerance, and disagreeing persons. *Child Development*, 75, 687-703.
- Wells, G. L., Lindsay, R. C., & Ferguson, T. J. (1979). Accuracy, confidence, and juror perceptions in eyewitness identification. *Journal of Applied Psychology*, 64, 440-448.
- West, P. M., & Broniarczyk, S. M. (1998). Integrating multiple opinions: The role of aspiration level on consumer response to critic consensus. *Journal of Consumer Research*, 25, 38-51.