Does informational independence always matter? Children believe small group discussion is more accurate than ten times as many independent informants

Emory Richardson (emory.richardson@yale.edu), Frank Keil (frank.keil@yale.edu)
Yale Department of Psychology, 2 Hillhouse Ave. New Haven, CT 06511 USA

Abstract

Learners faced with competing statements that each have support from multiple sources must decide whom to trust. Lacking firsthand knowledge, they frequently trust the majority. Yet, majorities can be misleading if most members are relying on hearsay from just a few members with firsthand knowledge. Thus, past work has emphasized the importance of informational independence when deciding whom to trust, showing that children and adults do consider informational independence important in certain contexts. However, because informational independence precludes group deliberation, we ask whether children make the reverse inference and devalue informational independence when facing a problem that could benefit from deliberation. In two studies, children and adults ignore informational independence when attempting to answer abstract reasoning questions. However, for a question type for which deliberative reasoning would be of doubtful benefit, children and adults seek advice from multiple independent sources rather than a deliberative group.

Keywords: group reasoning, trust in testimony, development, wisdom of crowds, cooperative learning

Introduction

In a famous demonstration, Galton (1907) showed that averaging 787 fair-goers' estimates of a cow's weight produced the exact weight in pounds. Averaging is an initially surprising means of extracting an accurate judgment from a large number of people, but the problem of managing information from multiple sources is ubiquitous in social groups. We rely daily on information many degrees removed from our own direct experience, and construct serviceable representations of reality from these indirect, incomplete, and error-prone representations, which are conveyed to us by agents in no better circumstances than ourselves. How do we do this? Here, we examine children and adults' intuitions about the value of independent testimony versus group discussion as two methods of managing collective capacities. We argue for an early developing recognition that deliberative reasoning is more effective in groups, regardless of the value of independent testimony in other domains.

The kind of question at hand may be important to deciding how to make use of collective intelligence. First, one might consider whether most people could provide a reasonably accurate answer, either because (A) everyone is approximately equally competent as a problem-solver or (B) the answer has already been figured out and is now common knowledge. If the competence landscape is fairly flat, one might do well to 'poll' a large number of informants and trust the majority judgement. Condorcet (1785) demonstrated that if independent voters each have a greater than 50% chance of correctly choosing between two options, the majority decision is more likely than the minority to identify the correct option; moreover, as the absolute number of independent voters increases, the majority approaches perfect accuracy. However, if the competence landscape is rugged — some people know essentially nothing, while others are domain experts — both majoritarian rule and the wisdom-of-crowds approach can harm accuracy: depending on the proportion of experts in the crowd of informants and the number of experts and cost function assigned to errors, the crowd may drown out accurate answers (Laan, Madirolas, de Polavieja, 2017).

Second, one might consider whether cooperative problem-solving could increase the odds answering correctly. Small group discussions frequently outperform not only the *average* individual, but also the *best* individuals on a variety of inductive reasoning tasks (Laughlin et al., 2006; Laughlin, 2011; Trouche, Sander, & Mercier, 2014), concept learning in undergraduate genetics courses (Smith et al., 2009), lie detection (Klein & Epley, 2015), and numerical estimation (Navajas et al., 2018). Of particular interest is the advantage in reasoning tasks. Deliberative reasoning may be most naturally deployed in service of argumentation, and function most effectively in interpersonal contexts (Mercier & Sperber 2011; Mercier, 2016). One reason for the group advantage may the ability to effectively handle greater amounts of evidence (Laughlin, Bonner, & Altermatt, 1998). By easing the cognitive load on individuals, groups allow members to monitor each others' reasoning for mistakes (Kirschner, Paas, & Kirschner, 2009a; 2009b). Increased processing capacity and the potential for error-correction may be particularly beneficial when some answers are demonstrably correct or incorrect (Laughlin & Ellis, 1985).

In short, effective social learning in multi-agent contexts requires some understanding of how information is distributed across individual agents and how different methods of aggregating collective

©2020 The Author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

knowledge and collective capacities influence the reliability of a response.

A variety of strategies support learning from others. For example, children are particularly likely to trust those who have been accurate in the past (Koenig & Harris, 2005; Birch, Vauthier & Bloom, 2008), and there is some evidence of a "copy-the-successful" rule in other species as well, including chimps (Menzel, 1974), bats (Wilkinson, 1992), and guppies (Lachland, Crooks, & Laland, 1998). Conforming to the majority is also a common learning heuristic in many species, including humans (Claidière & Whiten, 2012; Haun, Rekkers, & Tomasello, 2012). The majority bias is of particular interest because its value is contingent on the probability that the majority response is correct, or at least adaptive. When the majority is correct, the majority bias can be an effective way to reduce learning costs; when the majority is mistaken, trusting the majority can be far more costly than individual learning, depending on the consequences of adopting a maladaptive response. Thus, one should expect social learners to trust majorities selectively.

Children's trust in majorities is selective in a variety of ways. As noted earlier, the kind of question at hand may influence decisions about how to make use of collective intelligence. Conventions like object labels are intrinsically majoritarian, and 4-year-olds explicitly favor majority testimony for novel object labels (Corriveau, Fusaro, & Harris, 2009). Yet, when given reason to believe that a dissenter has privileged epistemic status with regard to a label — as might an artist for her art — children trust a single dissenter's label over a label provided by three other observers (Einav, 2014). Likewise, if a minority and majority *each demonstrate accuracy* by successfully manipulating a puzzle or tool, 4-year-olds endorse the minority as frequently as the majority (Hu et al., 2013; Burdett et al., 2016). These studies suggest that children do evaluate the competence of the majority and minority relative to the question at hand.

One dimension by which adults evaluate majorities may have a protracted developmental trajectory: the extent to which members themselves rely on their own judgement or are influenced by others. Hu et al., (2015) presented children with two boxes that each contained a different prize (different treats in one trial, different toys in another trial), and two sets of four informants. In one set, each member looked in each box, and so had "direct knowledge" of their contents; and each endorsed the same box. The other set endorsed the opposite box — but only one member had direct knowledge; the other three explicitly based their endorsement the fourth's testimony, without looking themselves. Four-year-olds preferred the endorsement based on direct testimony; yet, when two more informants were added to the "indirect knowledge" set in a second study, four-year-olds showed no preference. This suggests that while children favor direct knowledge, reliance on indirect knowledge does not discredit a majority in their eyes. Using a similar method, but an eyewitness memory paradigm, Aboody, Yousif, Sheskin & Keil (2019) more systematically varied the number of informants with direct and indirect knowledge across three experiments, and report that only by age 6 do children prefer a smaller number of informants with direct knowledge over a larger number with indirect knowledge. A more subtle form of independence prevents the informants from sharing any information at all — ensuring that they cannot mutually influence one another. Einav (2018) presented children with testimony about trivia, and found that only by age 8-9 did children endorse the testimony of three individuals independently recalling the same answer (without looking at each other's answers), with 5-year-olds actually preferring the testimony of the "non-independent" set in which two informants ostentatiously looked at the third's answer before answering themselves. Thus, not only were younger children not concerned that only one of the three in the non-independent consensus seemed to know the answer, but they may have interpreted the two informants "copying" as a cue to the first informant's expertise.

Both direct knowledge and "informational independence" can be important cues in some contexts. Direct knowledge may be valuable because independent responses can prevent the information cascades in collective judgements that reduce their reliability. Informational independence is also a key requirement for both the Condorcet jury theorem and the traditional interpretation of the wisdom of crowds. Because people tend to adjust their own responses to match others', collective accuracy may suffer if respondents are allowed to share information (Tversky & Kahneman, 1974; Epley & Gilovich, 2001; Larrick, Mannes, & Soll, 2012).

Nevertheless, while a strict insistence on direct knowledge or complete informational independence may be appropriate in "local" contexts such as eyewitness memory or judgements of everyday experiences, it is a poor general principle when there is a substantial division of cognitive labor. Popular belief in scientific theories is based on a few standard textbooks assigned to every high-schooler rather than direct knowledge of the evidence. It would seem silly to reject a scientific theory because scientists conferred together to develop it. Yet, insistence on informational independence as a prerequisite for trust even rules out *discussion* as a means of problem-solving. As noted earlier, small group discussions outperform individuals in a variety of reasoning tasks. Moreover, when the task has a demonstrably correct answer, one competent member is sufficient to predict an accurate collective decision, even if all other members initially agree on a wrong answer — given that the other members are sufficiently competent to understand the demonstration, "truth wins" (Laughlin & Ellis, 1986). As Laughlin & Ellis (1986) put it, "mathematics is the preeminent domain of demonstrability", but the "truth wins" also predicts the collective response to logical insight problems like the Tower of Hanoi and the Wason Card Selection task. Thus, a problem with a "demonstrably correct" answer simply needs to have an answer that clearly satisfies a set of constraints better than any alternative, given that all understand the constraints.

In the present studies, we chose a set of constraint-satisfaction problems that would challenge adults' capacities, but still be understandable to children. Because reasoning problems of this sort have a demonstrably correct solution, but the correct solution is difficult to identify among many possible alternatives, they are particularly likely to benefit from group error-correction (Laughlin, 2011). In contrast, informational independence may be important when correct answers are not demonstrable. A paradigm case is intuitions about population preferences. Without having ever polled compatriots, most of us have a sense of what our culture likes and dislikes; discussion may sway our intuitions one way or the other, but is unlikely to produce a "proof" of the correct answer. However, because past work has reported that children's grasp on the value of informational independence develops between the ages of 6 and 9 (Einav, 2018; Aboody et al., 2019), we predicted an overall bias for group discussion over independent testimony among younger children, with a question-specific preference for group discussion only emerging among older children.

Study 1

In Study 1, participants were asked which of two forms of advice would be more helpful for questions of two kinds: '*Reasoning*' and '*Population Preference*' (hereafter '*Popularity*'). The two forms of advice pitted the value of discussion with a group ('*Talking Together*') against value of independent testimony from each member of a crowd ('*Answering Alone*'). If participants recognize the specific value of group discussion, they should favor *Talking Together* for *Reasoning* more than for *Popularity*. If they value independent testimony over the benefits of group discussion regardless of question type, they should favor *Answering Alone* for both kinds of questions. We predicted that older children (9-10) and adults would favor *Talking Together* for *Reasoning* and *Answering Alone* for *Popularity*, while younger children (7-8) would favor group discussion for both kinds of questions. The pre-registration, data, and materials are available on the OSF.

Method

Participants. We recruited 40 adults through MTurk, as well as 80 children (40 Younger, M=8.01, SD=.56; 40 Older, M=9.92, SD=.56; 39 girls). Children participated through an online platform for developmental research that allows researchers to video chat with families using pictures and videos on slides (Sheskin & Keil, 2018). Sample size was chosen based on the estimated effect size from pilot results.

Materials. We asked eight test questions, four from each of two question types: '*Reasoning*' and '*Popularity*'. The *Reasoning* questions were chosen to be simple enough to explain to children, but challenging enough that adults would also be forced to think carefully to answer correctly. Questions were presented from the perspective of a protagonist (Jack). (1) A 4x4 Sudoku puzzle adapted for children. (2) A vehicle routing problem that required a MarioKart character to find the shortest road through all the treasures on a map without taking "two in a row that are the same color, or two in a row that are the same shape". (3) A single-heap, non-misère game of Nim ("Each side takes turns picking up pencils. Each turn, you can pick up either 1, 2, or 3 pencils. The winner is the person who picks up the last pencil. There are 5 pencils left in this game; how many pencils should Jack pick up?"). (4) An "impossible object" puzzle that requires the solver to remove an object from a bottle without breaking the object or the bottle. The *Popularity* questions concerned the most common subjective preferences in a population. (1) Whether pizza or hot dogs were more preferred by students in Jack's school. (2) What most people in the world say their favorite fruit is. (3) What most people in the world say their favorite day is. (4) What most people at Jack's school say their favorite day is. Questions were written to have approximately equal word counts (M_{Reas} = 73.75, M_{Pop}=67.75). Three counterbalances were created to vary the order of the questions — Forward, Reverse, and a Shuffle. Color coding of answer choice and left/right presentation were also counterbalanced between participants.

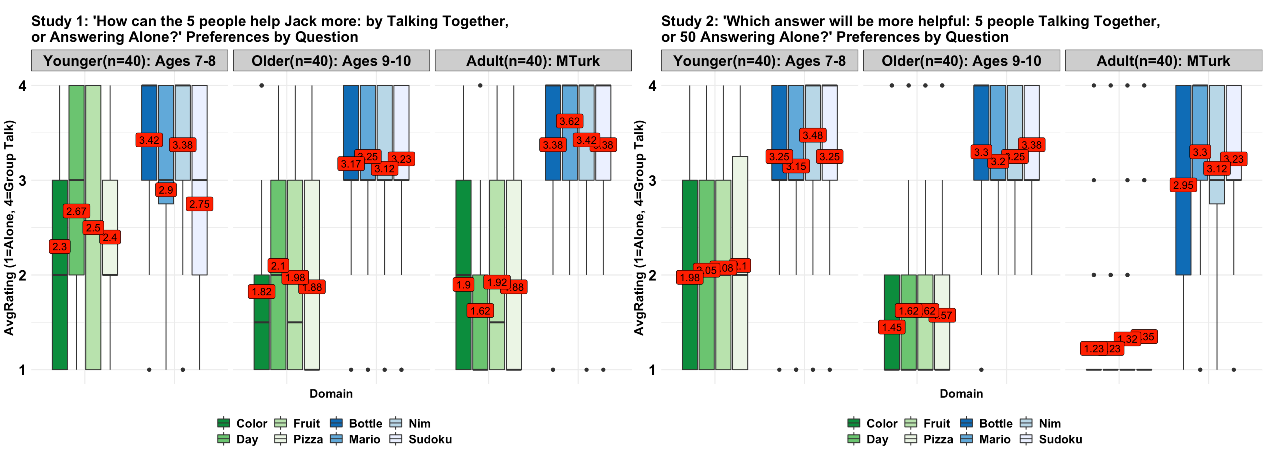
After the test questions, we asked two comprehension questions ("Comp_TT" and "Comp_AA") to test more explicitly whether participants were considering the effects of information sharing in a setting familiar to children. In these questions, Jack's teacher was giving a test to Jack's 5 informants, and participants were asked whether the 5 people should answer by Talking Together or by Answering Alone. In Comp_TT, the teacher wanted "the 5 people to get as many answers right as possible"; in Comp_AA, the teacher wanted to "find out which of the 5 people did their homework and which ones didn't". If children understand how discussion changes the informativeness of individual responses, they should recognize that Answering Alone is more informative to the teacher in Comp_AA. If they understand the benefits of discussion (or at least, information sharing), they should prefer Talking Together for Comp_TT.

Procedure. Children were introduced to the protagonist, Jack (a silhouette). They were told that Jack was unsure of the answers to the questions, and could ask five people for help. The five people could either help by *Talking Together* (giving Jack a single answer as a group), or by *Answering Alone* (each giving Jack their own answer after thinking about the question without consulting others). For each item, children first chose the better method and then were asked whether that method was "probably more helpful, or definitely more helpful"; adults used a 4-point scale directly. After answering the eight test items, participants were asked the two comprehension check questions (these were not counterbalanced: Comp_TT was always presented first). Two features of the procedure are important to keep in mind. First, participants could not evaluate the content of any answer to any question, because none was given; they were asked to choose a means of advice, not evaluate the quality of the advice itself. Secondly, they could not make judgements based on *degree* or *quality* of consensus — they only knew that the group would have to give one answer, while the crowd would have to give 5 independent answers which could differ or not.

Results and Discussion

For the primary test, the four responses within each question domain (Fig. 1) were averaged to create a single score for each domain. A repeated measures ANOVA revealed a significant effect of question *Type* ($F(1,117)=132.87$, $p<.001$, $\eta^2 = .523$) and an AgeGroup*Type interaction ($F(2,117)=7.83$, $p<.001$, $\eta^2 = .118$), and a marginal effect of AgeGroup ($F(2,117)=2.82$, $p=.064$, $\eta^2 = .046$). All age groups showed the predicted stronger preference for TalkingTogether for *Reasoning* questions, both as compared to *Popularity* questions (with Bonferroni corrections for multiple comparisons, Younger: $t(117) = 3.66$ $p=0.0057$, Older: $t(117)= 7.105$, $p<.0001$, Adult: $t(117) = 9.201$, $p<.0001$), and compared to chance responding (Younger: M=3.11, $t(39) = 7.42$, $p<.0001$, Older: M=3.19, $t(39) = 8.53$, $p<.0001$, Adult: M=3.45, $t(39) = 11.237$, $p<.0001$). However, Younger children were more similar to the adult pattern than we had predicted; while both Older children and Adults favored Answering Alone for *Popularity* questions, Younger children's answers for *Popularity* did not differ significantly from chance, thus showing a more mature sensitivity to question *Type* rather than the predicted general bias towards Talking Together (Younger: M=2.46, $t(39) = -0.232$, p=n.s., Older: M=1.94, $t(39) = -4.076$, $p<.001$, Adult: M=1.83, $t(39) = -5.24$, $p<.0001$). Moreover, children's responses to the comprehension questions suggest that even the youngest were able to choose a method of responding consistent with what the teacher wanted to learn about the students (Comp_AA: M_{Young}= 65%, $p=.04$, M_{Old}= 87.5%, $p<.0001$, M_{Adult}= 92.5%, $p<.0001$, Comp_TT: M_{Young}= 70%, $p=.008$, M_{Old}= 85%, $p<.0001$, M_{Adult}= 87.5%, $p<.0001$).

These data suggest that from an early age, people have an intuitive sense of how to manage collective wisdom. Children as young as 7 recognize that the benefits of deliberative reasoning in groups outweigh the risks introduced by non-independent responding; however, their preference for group discussion is not blind. By age 8-9, children prefer independent responding in a domain in which such responses may better approximate the truth than group deliberation. The pattern of responses from 7-8-year-olds supports the predicted early preference for group discussion as a means of answering reasoning questions, but suggests that their understanding of independent responding may be more mature than we had predicted. Moreover, children's responses to the two comprehension



Figures 1 & 2. Box plots of average preference for group discussion or independent responding in Study 1 (Fig. 1, left) and Study 2 (Fig. 2, right), presented by question for visualization (all analyses were on the average for Question Type). Blues=*Reasoning*, Greens=*Popularity.* Red labels are means for each question.

questions at the end of the study suggest that by age 7-8, they have explicit knowledge of how informational independence impacts the nature of information.

Study 2

How robust is the preference for group deliberation for reasoning questions? Group discussion allows individuals to vet an answer, but in our paradigm, also conveys less raw information than individual responding: the group discussion produces one answer, while the crowd poll produces five. In some cases more raw information may be desirable — for example, polling a larger crowd is more likely to accurately identify population preferences. Indeed, Condorcet's theorem suggests that group accuracy increases monotonically with group size (but see Kao & Couzin, 2014). However, increasing the number of informants is unlikely to improve accuracy for reasoning questions. First, increasing group size may quickly make effective deliberation unfeasible. Secondly, if individuals are typically unable to answer questions alone, polling a very large crowd is at best inefficient — while it may increase the odds of including an individual who *can* answer the question alone, it won't *identify* that individual in the crowd of wrong answers. In other words, a small group discussion may be a more efficient means of answering a reasoning question than the individual responses of even a very large crowd.

Thus, Study 2 made one change to Study 1: instead of a choice between asking five people to Talk Together or Answer Alone, the choice was between five people Talking Together or *fifty* people Answering Alone.

Our predictions for Study 2 were similar to those for Study 1. We predicted that adults and older children would continue to favor group deliberation for *Reasoning* questions, and crowdsourcing for *Popularity* questions. We predicted that younger children would continue to distinguish between *Reasoning* and *Popularity* questions, but saw two plausible alternatives for absolute judgements. First, younger children could show the mature pattern. Second, younger children's preference for group deliberation could be attenuated by a "more is better" bias for Reasoning items as well. The pre-registration and power analysis, and data and materials are available on the OSF.

Method

Participants. We recruited 40 adults through MTurk, as well as 80 children (40 Younger, M=8.01, SD=.56; 40 Older, M=9.92, SD=.56; 39 girls). As in Study 1, children participated through an online platform for developmental research that allows researchers to video chat with families using pictures and videos on slides (Sheskin & Keil, 2018). One additional child was excluded and replaced because the family lost internet connection partway through the study and could not rejoin.

Materials & Procedure. The materials and procedure were identical to Study 1, but the pictures now displayed fifty cartoon icons for *Answering Alone* instead of five.

Results and Discussion

For the primary test, the four responses for each question Type (Fig. 2) were averaged to create a single score for each Type. A repeated measures ANOVA revealed a significant effect of question Type ($F(1,117)=376.88$, $p<.001$, $\eta^2 = .763$) and AgeGroup ($F(2,117)=9.63$, $p<.001$ $\eta^2 = .141$)and an AgeGroup*Type interaction ($F(2,117)=5.39$, $p<.001$, $\eta^2 = .084$). The results were stronger than we expected. Even the youngest children showed the predicted stronger preference for *Talking Together* for *Reasoning* questions, both as compared to *Popularity* questions (with Bonferroni corrections for multiple comparisons, Younger: $t(117) = 8.60$ $p<.0001$, Older: $t(117)= 11.97$, $p<.0001$, Adult: $t(117) = 13.06$, $p<.0001$), and compared to chance responding (Younger: M=3.28, $t(39) = 9.29$, $p<.0001$, Older: M=3.28, $t(39) = 11.75$, $p<.0001$, Adult: M=3.15, $t(39) = 6.37$, $p<.0001$). As in Study 1, adults and older children favored *Answering Alone* for *Popularity* questions, but unlike Study 1, younger children also favored *Answering Alone* for *Popularity* questions (Younger: M=2.05, $t(39) = -3.38$, $p=.0017$, Older: M=1.57, $t(39) = -8.52$, $p<.0001$, Adult: M=1.28, $t(39) = -11.75$, $p<.0001$).

By age seven, children in our studies believed that a small group discussion would provide a more helpful answer than hearing the responses from even ten times as many individual informants. Yet, the youngest children also recognized that the responses of 50 individual informants would be more helpful to someone asking about population preferences. These results suggest that intuitions about how to effectively aggregate information from multiple sources develop early and in a context-specific manner; in particular, the recognition that group discussions are a more effective means of reasoning than individual effort or even asking ten times as many individuals is striking.

General Discussion

Informational independence is a potentially useful litmus test for the value of informant testimony in certain contexts (e.g., Yousif, Aboody, & Keil, 2019), but the division of cognitive labor that scaffolds human knowledge would be impossible if we were unable to trust the conclusions of collaborating groups. Our studies suggest that while children recognize the importance of informational independence in appropriate contexts, they also recognize that people reason more effectively in groups. When careful reasoning could produce a demonstrably correct answer to a question, children as young as seven believe that hearing from a small discussion group will be more informative than hearing from ten times the number of individuals.

These experiments suggest that children have abstract intuitions about how to manage collective wisdom; future work will examine to what extent children are able to put these intuitions into practice. Rekers, Haun, & Tomasello (2011) report that three year old children (but not chimpanzees) recruit partners for a collaboration even when able to complete a task alone; but recruiting group for deliberation or polling a large crowd require considerably more coordination and effort than recruiting a single partner. A fruitful approach might examine the relative costs of implementing each strategy; it seems likely that children apply a mixed strategy, wherein children "discuss" with an initial partner(s), and for more difficult problems continue to iteratively recruit more partners until they either solve the problem or give up. Just as it would be inefficient to recruit an informant simply to hear an answer without asking for additional explanation, it may be impractical to postpone discussion until some quorum is reached, or end discussion if one of several group members exits the discussion.

The Reasoning and Population Preference questions differed in more than just how "demonstrably" better one answer was. For example, Population Preference questions were more likely to elicit an intuitive response than the Reasoning questions, regardless of whether participants recognized the potential for a demonstrably correct answer to the Reasoning questions. A strong intuitive response may have made Population Preference questions appear easier. Likewise, children in particular may have believed the Population Preference questions had more than one answer, despite instructions to the contrary. Both possibilities are being addressed by a third experiment in which difficult perceptual discrimination tasks — which clearly have only one correct answer, but no intuitive response — are contrasted with easy versions of the same reasoning tasks. If the preference for group discussion is caused simply by the challenge of producing answers oneself, or by a desire for a single answer, then participants should prefer the group for challenging perceptual tasks.

Past work has suggested that though children begin to value direct knowledge from age 6 (Aboody et al., 2019), their appreciation of informational independence only develops by 8 or 9 (Einav, 2018). Our results are consistent with this trajectory, and suggest an explanation: children weigh the benefits of group discussion against the benefits of independent responding in a domain-specific fashion. By age 7, the group discussion is seen as more beneficial for solving reasoning problems than it is for discovering population preferences, and by age 9, independent responding is explicitly preferred for discovering population preferences.

Thus, we conjecture that children's mixed intuitions about the value of informational independence in contexts other than deliberative reasoning — such as certain forms of memory — may result from an early expectation that groups generally make cognitive processes more reliable, which is later refined into a more specific understanding of the benefits and risks of collective cognition for different kinds of question types.

We have argued that people may expect group discussion to produce more accurate answers to reasoning questions in particular because such questions characteristically have demonstrably correct answers, and so afford agents opportunity to produce and process potentially conclusive arguments. Though recent work has characterized testimony as more reliable when provided (A) independently or by (B) a greater number of direct sources, our work suggests that children and adults evaluate the importance of these factors in light of the question at hand. In a richly inductive domain, the testimony of a collaborating group outweighs even the testimony of ten times as many independent sources.

Acknowledgments

We thank the members of the Yale Cognition & Development lab for helpful feedback. This research was supported by NSF grant DRL 1561143 awarded to Frank C. Keil.

References

Aboody, R., Yousif, S. R., Sheskin, M., & Keil, F. C. (2019). Says who? Children consider informants' sources when deciding whom to believe. *Cognitive Development Society*, Louisville, KY.

Birch, S. A. J., Vauthier, S. A., & Bloom, P. (2008). Three- and four-year-olds spontaneously use others' past performance to guide their learning. *Cognition*, 107(3), 1018–1034.

Claidière, N., & Whiten, A. (2012). Integrating the study of conformity and culture in humans and nonhuman animals. *Psychological Bulletin*, *138*(1), 126–145.

Corriveau, K. H., Fusaro, M., & Harris, P. L. (2009). Going With the Flow: Preschoolers Prefer Nondissenters as Informants. *Psychological Science*, 20(3), 372–377.

Einav, S. (2014). Does the Majority Always Know Best? Young Children's Flexible Trust in Majority Opinion. *PLoS ONE,* 9(8), e104585.

Einav, S. (2018). Thinking for themselves? The effect of informant independence on children's endorsement of testimony from a consensus. *Social Development*, 27(1), 73–86.

Epley, N., & Gilovich, T. (2001). Putting adjustment back in the anchoring and adjustment heuristic: Differential processing of self-generated and experimenter-provided anchors. *Psychological Science*, 12(5), 391-396.

Galton, F. (1907). Vox Populi. *Nature* 75, 450–451.

Condorcet, M. D. (1785). Essay on the Application of Analysis to the Probability of Majority Decisions. *Paris: Imprimerie Royale*.

Haun, D. B. M., Rekers, Y., & Tomasello, M. (2012). Majority-Biased Transmission in Chimpanzees and Human Children, but Not Orangutans. *Current Biology*, 22(8), 727–731.

Hu, J., Whalen, A., Buchsbaum, D., Griffiths, T., & Xu, F. (2015). Can Children Balance the Size of a Majority with the Quality of their Information? *Proceedings of the Annual Meeting of the Cognitive Science Society*, *35*, 6.

Hu, J. C., Buchsbaum, D., Griffiths, T. L., & Xu, F. (2013). When does the majority rule? Preschoolers' trust in majority informants varies by task domain. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 6.

Kao, A. B., & Couzin, I. D. (2014). Decision accuracy in complex environments is often maximized by small group sizes. *Proceedings of the Royal Society B: Biological Sciences*, 281(1784), 20133305–20133305.

Kirschner, F., Paas, F., & Kirschner, P. A. (2009). A Cognitive Load Approach to Collaborative Learning: United Brains for Complex Tasks. *Educational Psychology Review*, 21(1), 31–42.

Kirschner, F., Paas, F., & Kirschner, P. A. (2009). Individual and group-based learning from complex cognitive tasks: Effects on retention and transfer efficiency. *Computers in Human Behavior*, 25(2), 306–314.

Klein, N., & Epley, N. (2015). Group discussion improves lie detection. *Proceedings of the National Academy of Sciences*, 112(24), 7460–7465.

Koenig, M. A., & Harris, P. L. (2005). Preschoolers Mistrust Ignorant and Inaccurate Speakers. *Child Development*, 76(6), 1261–1277.

Laan, A., Madirolas, G., & de Polavieja, G. G. (2017). Rescuing Collective Wisdom when the Average Group Opinion Is Wrong. *Frontiers in Robotics and AI*, 4.

Lachlan, R. F., Crooks, L., & Laland, K. N. (1998). Who follows whom? Shoaling preferences and social learning of foraging information in guppies. *Animal behaviour*, 56(1), 181-190.

Larrick, R. P., Mannes, A. E., & Soll, J. B. (2012). The Social Psychology of the Wisdom of Crowds. In J. I. Krueger (Ed.), *Frontiers of social psychology. Social judgment and decision making* (pp. 227–242). Psychology Press.

Laughlin, P. R. (2011). *Group Problem Solving*. Princeton University Press.

Laughlin, P. R., Bonner, B. L., & Altermatt, T. W. (1998). Collective versus individual induction with single versus multiple hypotheses. *Journal of Personality and Social Psychology*, 75(6), 1481–1489.

Laughlin, P. R., Hatch, E. C., Silver, J. S., & Boh, L. (2006). Groups perform better than the best individuals on letters-to-numbers problems: *Effects of group size. Journal of Personality and Social Psychology*, 90(4), 644–651.

Menzel Jr, E. W. (1974). A group of young chimpanzees in a one-acre field. In *Behavior of nonhuman primates* (Vol. 5, pp. 83-153). Elsevier.

Mercier, H. (2016). The Argumentative Theory: Predictions and Empirical Evidence. *Trends in Cognitive Sciences*, 20(9), 689–700.

Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, 34(2), 57–74.

Morgan, T. J. H., Laland, K. N., & Harris, P. L. (2015). The development of adaptive conformity in young children: Effects of uncertainty and consensus. *Developmental Science*, 18(4), 511–524.

Navajas, J., Niella, T., Garbulsky, G., Bahrami, B., & Sigman, M. (2018). Aggregated knowledge from a small number of debates outperforms the wisdom of large crowds. *Nature Human Behaviour*, 2(2), 126–132.

Rekers, Y., Haun, D. B. M., & Tomasello, M. (2011). Children, but Not Chimpanzees, Prefer to Collaborate. *Current Biology*, *21*(20), 1756–1758.

Sheskin, M., & Keil, F. (2018). TheChildLab.com A Video Chat Platform for Developmental Research. *PsyArXiv*.

Smith, M. K., Wood, W. B., Adams, W. K., Wieman, C., Knight, J. K., Guild, N., & Su, T. T. (2009). Why Peer Discussion Improves Student Performance on In-Class Concept Questions. *Science*, 323, 4.

Trouche, E., Sander, E., & Mercier, H. (2014). Arguments, more than confidence, explain the good performance of reasoning groups. *Journal of Experimental Psychology: General*, 143(5), 1958–1971.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty- Heuristics and biases. *Science*, 185(4157), 1124–1131.

Wilkinson, G. S. (1992). Communal nursing in the evening bat, Nycticeius humeralis. *Behavioral Ecology and Sociobiology*, 31(4), 225-235.

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.