Children learn the meaning of ambiguous evidence from third-party belief revision

Evan Orticio (eorticio@gmail.com)
Shihan Gao (eleanor_gao@berkeley.edu)
Celeste Kidd (celestekidd@berkeley.edu)
Department of Psychology,
University of California, Berkeley

Abstract

A person who changes their mind signals that they have encountered new information that prompted their belief shift. Can children use their developing understanding of third-party belief revision to take advantage of this signal for their own learning? Children (5;0-9;11 years) played a Whodunit-style game in which detectives updated their beliefs in response to different clues. The clues in isolation were meaningless to participants. In simple cases, children accurately inferred the meaning of the clues based on how they changed others' beliefs. With age, children more readily integrated changes in agents' certainty to guide these inferences. These findings suggest that children can draw reverse inferences about evidence by leveraging a causal understanding of how it impacts an agent's beliefs. Thus, children may learn world knowledge indirectly by observing its effects on others' minds.

Keywords: social inference, theory of mind, belief updating, social learning

Introduction

Cues from the social world help to constrain the learning space amidst a "blooming, buzzing confusion" of information. We use pedagogical cues to guide inductive inference (Butler & Markman, 2012), explore efficiently (Bonawitz et al., 2011), and access opaque cultural knowledge (Csibra & Gergely, 2009)—starting in early childhood but continuing throughout our lives. We leverage informant cues, like confidence, competence, and prior accuracy, to decide whom to trust and whom to fact-check (Harris et al., 2018; Orticio, Meyer & Kidd, 2024). The capacity to learn from testimony opens up a world of knowledge about topics for which data is inaccessible to individuals (Perfors & Navarro, 2019). In addition, we can learn from collective-level social cues by reasoning about the distribution of beliefs among different people. Children temper their beliefs and explore more in the context of disagreement (Langenoff, Srinivasan & Engelmann, 2024), and take consensus as a cue to reliability (Fusaro & Harris, 2008; Aboody et al., 2022). The social world provides new windows into knowledge and guides efficient information search.

Learning from when others change their minds

Children may also learn from mental state dynamics, like an agent's changes in beliefs. Consider a child who secretly eats a bunch of candy in the morning without her father suspecting a thing. Later, the child is acting hyperactive. Her dad comments on her unusually high energy and accuses her of sneaking some candy earlier that day. The child can learn two things from her parent's change in belief. First, she has been caught. Second, she can learn that candy can cause a sugar rush. By observing how the evidence (their hyperactivity) caused her dad to increase his belief that she had eaten candy, the child can infer the knowledge that candy causes hyperactivity. In other words, she can learn a new way to interpret evidence (in this case, her hyperactivity), when she may have been unaware that the evidence meant anything in the first place.

Changes of beliefs are potent signals for learning. Belief change inherently indicates that the information that triggers it was influential. Beliefs can be quite stubborn (Martí et al., 2018; Wade & Kidd, 2019), so a change in belief suggests the evidence that prompted it was especially compelling. Focusing on what data prompts belief revision can help a learner focus on just the information that is most valuable, which is important given the fact that people form beliefs from minimal data (Klein & O'Brien, 2018). Belief change is also sensitive to an agent's priors, so identifying changes in belief can be a window into the agent's prior knowledge.

While attending to belief changes may require all of the necessary computational costs associated with metacognitive tasks in general, we know that inferring mental states is something people do naturally via a principle of rational action (Baker et al., 2017; Jara-Ettinger et al., 2015; Jara-Ettinger et al., 2016). For example, as early as 10 months of age, infants infer the value of goals from the costs of achieving them, expecting agents to prefer rewards for which they were willing to jump farther or climb higher to reach (Liu et al., 2017). This model of other minds can also be applied to epistemic actions. Four-year-olds infer an agent's epistemic states from their willingness to incur costs to gain information (Aboody, Zhou & Jara-Ettinger, 2021). In this study, children judged that an agent who declined low-cost information must have already known it, and an agent who incurred a greater cost to gain information must have really wanted to know it. These results suggest that children are able to reason about the value that a specific agent puts on information, which indexes an agent's level of uncertainty in a given belief. Thus, it is plausible that children could possess the metacognitive and mentalizing skills required to interpret evidence based on how it elicits nuanced changes in another agent's belief.

Children's understanding of the dynamics of third-party belief updating

Children have principled expectations about third-party belief dynamics that support such inferences. Five-year-olds demonstrate intuitions about the typical trajectories of epistemic states, expecting that gaining knowledge is more common than losing knowledge (Briscoe, Zhang & Jara-Ettinger, 2024). Children rationally integrate an agent's observed data, the sampling process of that data, and their prior beliefs when predicting whether the agent will change their mind (Magid et al., 2016). Children also infer that a speaker is knowledgeable when they have causal influence over listeners (Chuey, Sparks & Gweon, 2023). Finally, children distinguish between an agent who makes accurate observations and an agent who makes accurate predictions, and infer that only the latter was knowledgeable (Aboody, Huey & Jara-Ettinger, 2022). These findings underscore a causal understanding of the interactions between access to information, mental states, and actions.

While it is clear that children's inferences are supported by a sophisticated causal model of mental state changes, prior work has largely focused on how these expectations guide inferences about agents rather than inferences about the external world. One recent study examined how children infer what happened when two agents disagree (Amemiya, Heyman & Gerstenberg, 2024). The study found that older children inferred that evidence was ambiguous when the agents disagreed on its interpretation. Thus, children can use other agents' reactions to evidence to infer its qualities, despite lacking firsthand access to the evidence itself. However, the study examined disagreement between two parties, which likely involves distinct computations from those involved in interpreting belief change within a single agent. Moreover, the form of the evidence in this study was a speaker's utterance, so participants were making inferences about communicative intent. It remains unclear whether children can use a similar mechanism to ascertain objective, yet unobserved, facts about the world from another agent's mental states.

Present Study

We test the possibility that children can learn from third-party belief revision. We created a Whodunit-style game in which children watched detectives solve a case. Detectives stated their beliefs about which suspect was guilty of a crime before and after discovering clues on the crime scene. We examine children's ability to infer the meaning of a given clue (i.e., which suspect was associated with the clue) on the basis of how the detectives' suspicions changed. For example, if a detective discovers a hat on the crime scene and increases their belief that Martin is a thief, can children infer that the hat belongs to Martin? This would provide evidence that children can infer knowledge indirectly by monitoring changes in another agent's beliefs, and use that knowledge to interpret ambiguous evidence. In an experiment with children aged 5-9, we test two kinds of evidence (positive vs. negative) and two kinds of belief change (changing discrete belief states vs. changing certainty) across four trial types.

Method

Participants

Sixty 5- to 9-year-old children (M = 7.38 years, SD = 1.50, 38 girls) were tested on campus (n = 6) or in parks (n = 54) in the San Francisco Bay Area. All children provided verbal assent before participating, and those aged 7 and older also signed assent forms. Children received a small toy valued at $1-2 for their participation, and parents who came to campus were compensated for travel. Nine additional children were excluded from the analysis due to experimenter error or because they observed another child participating.

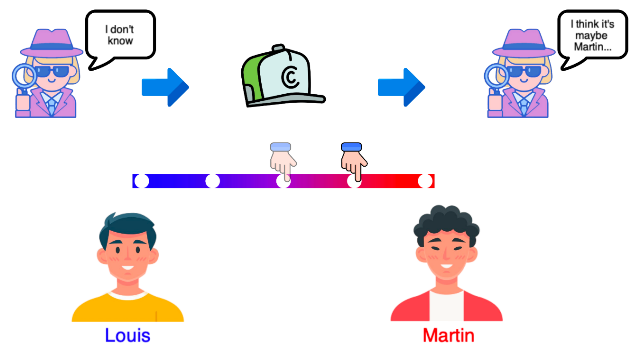

Figure 1: Example adopt-belief trial. Once the detective sees the clue (hat) and changes her mind about who committed the crime, the pointer moves to represent her final belief on the visual scale.

Materials and Procedure

Training Phase Children were introduced to a computer game created in PsychoPy. They were told they needed to help a group of detectives figure out who among two suspects was a thief. Detectives shared their beliefs about which suspect was the thief using one of the following phrases on a 5-point scale: "I really think it's Suspect A!" (1), "I think it's maybe Suspect A…" (2), "I don't know" (3), "I think it's maybe Suspect B…" (4), and "I really think it's Suspect B!" (5). The experimenter narrated these beliefs with prosodic cues consistent with the level of certainty of the belief. To scaffold children's understanding, the belief was also visually represented using a pointer on a 5-point scale. Children were shown three examples of detectives' beliefs and how they were represented on the scale. Then, they completed one comprehension check question in which they had to map a new detective's belief onto the 5-point visual scale. If the child selected the wrong corresponding location on the scale, the experimenter corrected them and explained the mapping before proceeding. Finally, children were shown that detectives could change their minds, and that changes in

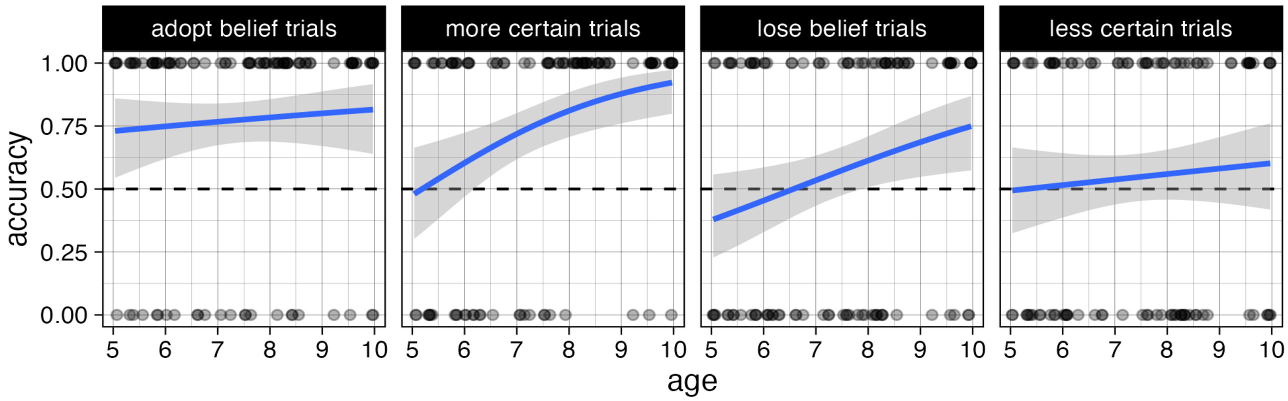

Figure 2: Children successfully inferred the meaning of evidence in adopt-belief trials and more-certain trials. Performance on more-certain trials and lose-belief trials improved with age. Dots represent individual trials. Logistic regression lines are shaded with 95% CIs.

belief would be represented visually by an animated pointer moving from their initial belief to their final belief.

Test Phase Children completed two blocks of 4 trials. Each block contained one of four trial types corresponding to four forms of belief change: adopting a belief, becoming more certain in a belief, losing a belief, and becoming less certain in a belief. Each trial type represented a movement to an adjacent point on the 5-point scale. In **adopt-belief trials**, the detective changed her belief from neutral ("I don't know") to a moderate belief about one suspect ("I think it's maybe X…") after observing the new clue. In **more-certain trials**, the detective changed from a moderate belief about a suspect ("I think it's maybe X…") to a stronger belief about that suspect ("I really think it's X!"). In **lose-belief trials**, the detective changed from a moderate belief about a suspect ("I think it's maybe X…") to a neutral belief ("I don't know"). In **less-certain trials**, the detective changed her belief from a strong belief about one suspect ("I really think it's X!") to a more moderate belief about that suspect ("I think it's maybe X…"). Trial order was randomized within each block, and suspects were counterbalanced between participants.

On each test trial, children saw an independent detective state her initial belief of which of the two suspects was the thief, discover a new clue on the crime scene (e.g., a hat), and then state her final belief (see Figure 1). Participants were told that the detective knew which of the two suspects was associated with the clue (e.g., who usually wears this hat), and that this made her change her mind a little bit. Then, participants were asked, "Who do you think was wearing the [clue]?". No feedback was given during test trials. Note that children were asked about which suspect was associated with the clue, and not which suspect was the thief. Thus, this measure reflects their ability to make inferences about an otherwise ambiguous clue on the basis of the detective's implicit reasoning.

Participants had to note the *direction* of the detective's belief change upon seeing the clue in order to correctly infer which suspect was wearing the given clue. In adopt-belief trials and more-certain trials, the correct answer is the suspect that the detective ultimately endorses after seeing the clue. However, in lose-belief trials and less-certain trials, the detective either abandons or reduces her belief in her initial suspect. In these cases, seeing the clue caused the detective's belief to move in the direction of the other suspect, so the other suspect must have been wearing the given clue.

Results

Analyses were run in R 4.3.1, using the *lme4* and *stats* packages. Seven participants failed the comprehension check question and were excluded. However, we ran an exploratory analysis to assess whether children's performance was impacted by their understanding of how verbal beliefs mapped onto the visual scale. We fit a mixed-effects logistic regression model predicting accuracy, with trial type and comprehension check performance as fixed effects and random intercepts for each participant. The model revealed no main effect of comprehension check performance and no significant interactions with any of the trial types (all p's > .05), suggesting that children's performance on the task did not significantly depend on visual scaffolding.

Children infer the meaning of positive evidence from third-party belief change

We ran four binomial tests comparing overall performance to chance (50%) to assess performance in each trial type individually. Children's accuracy in selecting the correct suspect was significantly above chance in adopt-belief trials (M = 77.36%, 95% CI [68.21, 84.92], $p < .001$) and in more-certain trials (M = 73.58%, 95% CI [64.13, 81.68], $p < .001$). Thus, children successfully inferred an agent's knowledge from their change in belief in response to positive evidence. Overall accuracy was not significantly different from chance in lose-belief trials (M = 56.60%, 95% CI [46.63, 66.20], $p = .21$) or less-certain trials (M = 54.72%, 95% CI [44.75, 64.41], $p = .38$). This suggests that, on a group level, children had difficulty making clear inferences on the basis of negative evidence.

Inferences from third-party changes in certainty improve with age

To assess how this inferential ability changes across ages, we ran four mixed-effects logistic regression models using mean-centered age to predict accuracy in each of the four trial types (see Figure 2). Age was a significant predictor of accuracy on more-certain trials (β = 1.16, p = .016). Older children were more likely than younger children to select the correct suspect on the basis of the detective's increase in certainty upon seeing the clue. Also, age significantly predicted performance on lose-belief trials (β = 0.57, p = .029). Older children were more likely than younger children to successfully infer the meaning of evidence that causes a detective to abandon an initial belief and become unsure. Age was not a significant predictor in adopt-belief trials (β = 0.16, p = .52) or less-certain trials (β = 0.14, p = .49).

We conducted an exploratory analysis to determine whether older children succeed on lose-belief trials, motivated by the significant age effect. A binomial test revealed that older children (greater than or equal to the median age of 7.61 years) performed significantly above chance in lose-belief trials (M = 68.52%, 95% CI [54.45, 80.48], p = .009).

Discussion

Children exploit social cues to orient themselves toward the most important, learnable information in their environment. Here, we investigated whether children can use third-party belief revision as such a cue. We find that children can infer the meaning of otherwise ambiguous evidence by observing how that evidence influences another agent's beliefs. This is consistent with an ability to use a mentalistic model of third-party belief updating to uncover an agent's world knowledge from the dynamics of their beliefs. In particular, success in our task likely involved (1) tracking the direction of change in an agent's belief from their verbal testimony, scaffolded by visual cues; (2) recognizing the causal relationship between the observation of a clue and the change in the agent's belief; and (3) reasoning backward to infer the agent's latent world knowledge about that clue.

The present research suggests that children take into account agents' belief updating process to support their inferences, rather than merely matching their static beliefs. It is possible in principle that children selected the suspect that the detective endorsed without reasoning about their belief change process at all. This belief-matching account is consistent with the fact that, on a group level, children only succeeded on adopt-belief and more-certain trials: the two trial types for which the correct answer also happens to be the suspect that the detective explicitly endorsed. However, several findings speak against this confounding account. First, the belief-matching account would predict a floor effect for lose-belief and less-certain trials, since the suspect that the detective endorsed is the incorrect answer. Instead, we observed chance-level performance even among the youngest participants, which may reflect difficulty with reasoning about negative evidence and integrating certainty.

Second, the belief-matching account would predict at-ceiling performance on more-certain trials, and not an age trend like we observed. If participants were merely matching the detective's stated belief, there is no reason to expect younger children to be performing at chance in more-certain trials but not in adopt-belief trials. The difficulty with more-certain trials is more likely attributable to the added complexity of integrating certainty into their model of the detective's beliefs.

Older children also correctly infer that evidence which makes an agent abandon an initial belief must support the opposite conclusion. Success on lose-belief trials is further evidence against the belief-matching account, because selecting the suspect that the detective endorsed on these trials would yield an incorrect answer. It should be noted that the separate analysis of older children was exploratory, so further work should verify children's ability to integrate negative evidence into these mentalistic inferences. However, the observed age effect is reasonable in light of the increased demands of lose-belief trials. By design, the detective endorses the incorrect suspect in the beginning of these trials. Thus, children may struggle to inhibit this more salient suspect in favor of the alternative suspect, who was never explicitly mentioned in the trial. Indeed, an additional exploratory analysis revealed that accuracy on lose-belief trials significantly increased with trial number (β = 0.30, p = .019), demonstrating that more experience with the task improved performance. Accuracy on the final lose-belief trial was above chance across our full age range (M = 66.04%, 95% CI [51.73, 78.48], p = .027). These exploratory findings suggest that task-specific demands, and not reasoning deficits per se, may explain younger children's chance-level performance.

Our findings imply that the ability to reason about changes in certainty develops later than changes in discrete belief states. Children in our study failed to make reliable inferences from an agent's increasing certainty until age 6 or 7. Younger children may have had difficulty identifying or distinguishing the levels of certainty, because they differed only slightly through the use of certain lexical items, like "maybe" vs. "really think", and prosodic cues. Nevertheless, there is evidence that children are sensitive to linguistic indicators of certainty by age 3 (Matsui et al., 2016), and confidence levels were simultaneously represented on a visual scale. Children may instead have trouble *reasoning* about third-party certainty (Pillow, 2008). Evidence suggests that the ability to reason about others' certainty is separate from first-person metacognition (Baer, Malik & Odic, 2021) and more protracted in development (Miosga et al., 2020). Our data are consistent with findings that young children can reason about how information-seeking relates to knowledge (vs. ignorance) by age 4 (Aboody, Zhou & Jara-Ettinger, 2021), but struggle to make the same mapping onto states of certainty until at least age 6 (Huang, Hu & Shao, 2019).

Children's difficulty making reliable inferences from negative evidence may reflect both task-specific limitations and broader conceptual challenges. In lose-belief and less-certain trials, the clue constituted negative evidence for the detective's initial belief. This complicates the reasoning process on several levels. For one, it depends more explicitly on a mutual exclusivity assumption: that every clue must pertain to one of the two suspects. Although we make this assumption explicit during training, it may not have been salient to children. Additionally, both trials are pragmatically strange because the clue does not cause the detective to switch fully to endorsing the other suspect. This was intentional in order to create stringent tests of mentalistic reasoning under conditions parallel to the positive evidence trials. However, it meant that the clue was indicative of the alternate suspect yet not strong enough evidence to explicitly endorse that suspect. It may have been difficult for children to accommodate this discrepancy, or to conceptualize weak evidence. These issues limit our conclusions, and motivate future work elucidating children's understanding of partial or inconclusive evidence.

The exact content of children's inferences remains unclear. Children's success in the task is notable given that the evidence was circumstantial and the detectives' reasoning was therefore abductive or underdetermined. Detectives could change their belief based on knowledge that a clue belonged to a given suspect, knowledge that a clue did not belong to the alternative suspect, inferences about the clue's owner based on the suspects' preferences (e.g., Louis hates green, so that green hat isn't his), or many other possibilities. Future work should test the content and flexibility of these inferences, for example by evaluating whether they're defeasible in light of additional information about the suspects or context. In addition, planned work investigates how children infer more abstract properties of the evidence, like evidential strength, and use those inferences to guide learning.

In sum, the present work highlights that children can use others' mental *processes*, on top of their mental states, as data for their own learning. This capacity facilitates efficient learning because instances of belief change are rare signals of influential information. It also may help resolve disagreements or improve persuasion. By mapping out a structured representation of another's knowledge and beliefs, children could learn to devise more effective interventions on that system and become stronger communicators (Ho, Saxe & Cushman, 2022). More broadly, sensitivity to others' belief dynamics reflects an important developmental shift in focus from outcome, i.e., the content of an agent's beliefs, to process, i.e., how those beliefs are formed (Butler, Gibbs & Tavassolie, 2020; Schleihauf et al., 2023). It is critical that children are equipped with the knowledge of *how* rational belief formation unfolds as they navigate increasingly complex social and informational ecosystems.

Acknowledgements

We are grateful to Leah Abazari, Nia Elrick, Sherry Huang, Elisa Matalon, and Emily Nightingale for their assistance with data collection.

References

Aboody, R., Huey, H., & Jara-Ettinger, J. (2022). Preschoolers decide who is knowledgeable, who to inform, and who to trust via a causal understanding of how knowledge relates to action. *Cognition*, *228*, 105212.

Aboody, R., Yousif, S. R., Sheskin, M., & Keil, F. C. (2022). Says who? Children consider informants' sources when deciding whom to believe. *Journal of experimental psychology: general*, *151*(10), 2481.

Aboody, R., Zhou, C., & Jara-Ettinger, J. (2021). In pursuit of knowledge: Preschoolers expect agents to weigh information gain and information cost when deciding whether to explore. *Child Development*, *92*(5), 1919-1931.

Amemiya, J., Heyman, G. D., & Gerstenberg, T. (2024). Children use disagreement to infer what happened. *Cognition*, *250*, 105836.

Baer, C., Malik, P., & Odic, D. (2021). Are children's judgments of another's accuracy linked to their metacognitive confidence judgments?. *Metacognition and Learning*, *16*(2), 485-516.

Baker, C. L., Jara-Ettinger, J., Saxe, R., & Tenenbaum, J. B. (2017). Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. *Nature Human Behaviour*, *1*(4), 0064.

Bonawitz, E., Shafto, P., Gweon, H., Goodman, N. D., Spelke, E., & Schulz, L. (2011). The double-edged sword of pedagogy: Instruction limits spontaneous exploration and discovery. *Cognition*, *120*(3), 322-330.

Briscoe, M., Zhang, R., & Jara-Ettinger, J. (2024). Children's Expectations About Epistemic Change. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 46).

Butler, L. P., Gibbs, H. M., & Tavassolie, N. S. (2020). Children's developing understanding that even reliable sources need to verify their claims. *Cognitive Development*, *54*, 100871.

Butler, L. P., & Markman, E. M. (2012). Preschoolers use intentional and pedagogical cues to guide inductive inferences and exploration. *Child Development*, *83*(4), 1416-1428.

Chuey, A., Sparks, R., & Gweon, H. (2023). Young children can identify knowledgeable speakers from their causal influence over listeners. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 45, No. 45).

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, *13*(4), 148-153.

Fusaro, M., & Harris, P. L. (2008). Children assess informant reliability using bystanders' non-verbal cues. *Developmental Science*, *11*(5), 771-777.

Harris, P. L., Koenig, M. A., Corriveau, K. H., & Jaswal, V. K. (2018). Cognitive foundations of learning from testimony. *Annual Review of Psychology*, *69*(1), 251-273.

Ho, M. K., Saxe, R., & Cushman, F. (2022). Planning with theory of mind. *Trends in Cognitive Sciences*, *26*(11), 959-971.

Huang, Z., Hu, Q., & Shao, Y. (2019). Understanding others' knowledge certainty from inference and information-seeking behaviors in children. *Developmental Psychology*, *55*(7), 1372.

Jara-Ettinger, J., Gweon, H., Schulz, L. E., & Tenenbaum, J. B. (2016). The naïve utility calculus: Computational principles underlying commonsense psychology. *Trends in Cognitive Sciences*, *20*(8), 589-604.

Jara-Ettinger, J., Gweon, H., Tenenbaum, J. B., & Schulz, L. E. (2015). Children's understanding of the costs and rewards underlying rational action. *Cognition*, *140*, 14-23.

Klein, N., & O'Brien, E. (2018). People use less information than they think to make up their minds. *Proceedings of the National Academy of Sciences*, *115*(52), 13222-13227.

Langenhoff, A. F., Srinivasan, M., & Engelmann, J. M. (2024). Disagreement reduces overconfidence and prompts exploration in young children. *Child Development*.

Liu, S., Ullman, T. D., Tenenbaum, J. B., & Spelke, E. S. (2017). Ten-month-old infants infer the value of goals from the costs of actions. *Science*, *358*(6366), 1038-1041.

Magid, R. W., Yan, P., Siegel, M. H., Tenenbaum, J. B., & Schulz, L. E. (2018). Changing minds: Children's inferences about third party belief revision. *Developmental Science*, *21*(2), e12553.

Martí, L., Mollica, F., Piantadosi, S., & Kidd, C. (2018). Certainty is primarily determined by past performance during concept learning. *Open Mind*, *2*(2), 47-60.

Matsui, T., Yamamoto, T., Miura, Y., & McCagg, P. (2016). Young children's early sensitivity to linguistic indications of speaker certainty in their selective word learning. *Lingua*, *175*, 83-96.

Miosga, N., Schultze, T., Schulz-Hardt, S., & Rakoczy, H. (2020). Selective social belief revision in young children. *Journal of Cognition and Development*, *21*(4), 513-533.

Orticio, E., Meyer, M., & Kidd, C. (2024). Exposure to detectable inaccuracies makes children more diligent fact-checkers of novel claims. *Nature Human Behaviour*, 1-8.

Perfors, A., & Navarro, D. J. (2019). Why do echo chambers form? The role of trust, population heterogeneity, and objective truth. In *CogSci* (pp. 918-923).

Pillow, B. H. (2008). Development of children's understanding of cognitive activities. *The Journal of Genetic Psychology*, *169*(4), 297-321.

Schleihauf, H., Zhang, Z., Gomez, A., & Engelmann, J. M. (2023). From outcome to process: A developmental shift in judgments of good reasoning. *Cognition*, *236*, 105425.

Wade, S., & Kidd, C. (2019). The role of prior knowledge and curiosity in learning. *Psychonomic Bulletin & Review*, *26*, 1377-1387.