

Children consider the probability of random success when evaluating knowledge

Rosie Aboody¹ (rosie.aboody@yale.edu), Stephanie Denison² (stephanie.denison@uwaterloo.ca), & Julian Jara-Ettinger¹ (julian.jara-ettinger@yale.edu)

¹Department of Psychology, Yale University. New Haven, CT 06520 USA

²Department of Psychology, University of Waterloo. Waterloo, Ontario, N2L 3G1

Abstract

To infer what others know, we must consider under what epistemic states their actions were both rational and probable. We test whether preschoolers can compare the probability of different actions (and outcomes) under different epistemic states—and use this to evaluate what others know. Specifically, four- to six-year-olds ($n=90$) were asked to help evaluate an agent's knowledge state by asking the agent to complete either an “undiagnostic” task (where success was assured), or a “diagnostic task” (where the probability of random success was low). By age six, children understood that the “diagnostic” task would more likely reveal the agent's knowledge state; four- and five-year-olds had no reliable preference, although children in all age groups understood that the “diagnostic” task was harder. These results suggest that, by the end of preschool, children understand how agents' epistemic states and environment jointly determine success—considering whether agents' actions imply knowledge, or just luck.

Keywords: social learning; knowledge; probability

Introduction

To navigate the social world, it is important to understand what others believe and know. Because we can never see into others' minds, we often must infer their epistemic states from their behavior. As adults, such inferences are commonplace: if your friend is hurrying to work but skips a shortcut, you might infer they don't know about it. And if a new lab-mate immediately presses the right buzzer to unlock the lab door—ignoring three identical but irrelevant buttons—you might infer that someone has already shown them around.

Despite their ubiquity, such epistemic inferences are far from straightforward. While it may be convenient to infer ignorance from failure (e.g., skipping a shortcut) and knowledge from success (e.g., opening a door), the relationship between epistemic states and actions is not always deterministic. For instance, if all four buzzers actually unlocked the lab door (or the correct buzzer was prominently labeled), your lab-mate's success might not clearly reveal how much they already knew, since they would have succeeded no matter what. And if the buzzer just broke, even someone who knows how to get in might not be able to open the door. Thus, to infer what others know, we cannot consider only the outcome of their actions—we must also consider the reasons behind them.

Recent research suggests that even preschoolers do not rely solely on action outcomes to infer what others know (Aboody et al. 2018, 2019; Einav & Robinson, 2011; Jara-Ettinger et al., 2017). For instance, four- and five-year-olds infer that an agent who refuses to pursue minimally costly information must have already known it, and an agent who decides to pursue high-cost information must have really wanted it (Aboody et al., 2021). This and related work (Jara-Ettinger et al., 2017) suggests that by the end of preschool, children rely on an expectation that agents maximize utilities (selecting actions that will yield the greatest rewards and incur the fewest costs) to infer others' epistemic states or motivations.

However, as the buzzer example illustrates, the relation between knowledge and action is mediated by the state of the world. To identify the reasons behind others' actions, we must consider whether these actions were truly diagnostic of their epistemic states. That is, to accurately infer what others know, children must be able to identify when an action is only likely given knowledge (e.g., picking the right buzzer when three don't work)—and understand that the exact same action could be ambiguous under different circumstances (e.g., if all four buzzers open the door).

It is possible that even young children consider others' actions in context, comparing the probability of agents' actions under different epistemic states to decide whether their behavior is more consistent with knowledge than ignorance. Children are sensitive to probability from infancy, distinguishing probable from improbable outcomes in the first years of life (Denison et al., 2013; Xu & Garcia, 2008; Gweon et al., 2010). Additionally, recent research into children's epistemic inferences suggests that preschoolers may consider the probability of a random success at least qualitatively: for instance, attributing knowledge to an agent who can predict an otherwise difficult-to-guess outcome, but not to an agent who observes the outcome and simply describes it (Aboody et al., 2018).

However, it is also possible that young children do not consider the probability of an agent's actions under different epistemic states when inferring knowledge. Although preschoolers are extremely attentive to the outcome of others' actions (preferring to learn from and endorse the testimony of agents who were previously accurate; Corriveau & Harris, 2009; Koenig et al., 2004; Koenig & Harris, 2005; Harris 2012), they do not always consider those outcomes in

context. For instance, preschoolers are not always sensitive to the reasons behind others' errors, trusting (Bridgers et al., 2016) or distrusting (Nurmsoo & Robinson, 2009) agents without distinguishing whether their past errors were justified. Further, it is unclear whether children fully understand what it means to be ignorant: preschoolers do not expect ignorant agents to search randomly between two locations (Ruffman 1996; Saxe, 2005; Chen et al., 2015; Friedman & Petrashek, 2009), and attribute greater expertise to those who confidently answer unknowable questions rather than those who correctly demur (Kominsky et al., 2016).

While a small number of studies have found that children are sensitive to probability when inferring desires (Diesendruck et al., 2015; Kushnir et al., 2012), developmental psychologists have long noticed asymmetries in children's understanding of mental states like goals or desires, and epistemic states like beliefs. From the first years of life, children are able to both represent and infer others' goals and desires (Gergely & Csibra, 2003; Liu et al., 2017; Woodward, 1998). However, it is not until age four or five that children can explicitly represent others' beliefs (for a review see Wellman et al., 2001). And even preschoolers who can represent epistemic states may not fully understand how to infer them: recent research suggests that an ability to infer epistemic states continues to develop between age four and six (Aboody et al., 2018, 2019; Wu & Schulz, 2018). Thus, it is unclear whether children can leverage their early-emerging understanding of probability to evaluate others' epistemic states, in addition to inferring their desires.

In the current study, we test whether four- to six-year-olds consider the probability of a chance success when evaluating others' epistemic states. Specifically, we test whether children understand that asking an agent to complete a "diagnostic" task (with only a 25% chance of random success) would better reveal their knowledge state, in contrast to an "undiagnostic" task (where success is assured). We focus on this age range because children's belief reasoning (Wellman et al., 2001; Wellman, 2014) and understanding of ignorance (Friedman & Petrashek, 2009; Ruffman, 1996) is still developing during the preschool years. Furthermore, recent work suggests that an ability to use probability to predict others' emotional reactions is still developing between age four and six (Doan et al., 2018).

Experiment

Method

The procedures, predictions, sample size, exclusion criteria, and analysis plan were pre-registered (see OSF page: https://osf.io/e6b5h/?view_only=13f04c99e81c4220b79d1d9a5bec650b).

Participants 90 four- to six-year-olds (mean age: 5.51 years, range: 3.96-6.9 years) participated. Eleven additional participants were recruited but not included in the study (see

Results). All participants completed the experiment via an online video-chat research platform.

Stimuli Stimuli consisted of a Powerpoint presentation, featuring a cartoon character of a girl, four blue boxes lined up on a blue background, and four green boxes lined up on a green background. Five of the boxes (four on one side, and one on the other) had a yellow marble hidden underneath; three of the boxes were empty (see Figure 1).

Procedure Figure 1 shows the experimental procedure. The experiment always began with eight boxes appearing on the screen. On the left were four blue boxes lined up on a blue background, and on the right were four green boxes lined up on a green background. The experimenter began by pointing out the boxes, saying, "Look! There are blue boxes on the blue side, and green boxes on the green side. Let's look under all of the boxes!" Starting on the blue side, the experimenter lifted each box one at a time to reveal its contents. Participants saw that every box on the blue side had a marble underneath (the "undiagnostic" side). The experimenter described each box's contents as they were revealed, saying, "Look, there's a marble under this box!" After lifting all of the boxes, the experimenter recapped by saying, "So, *all* of the boxes on the blue side have a marble underneath" (text italicized to mark words emphasized by the experimenter).

The experimenter then moved onto the green side, repeating the same procedure. Only one box on the green side had a yellow marble underneath (the marble was always under the third box), while the other boxes were empty (the "diagnostic" side). The experimenter described the box with the marble in the same way as before, and described the empty boxes by saying, "Look, there's nothing under this box." Finally, the experimenter recapped by saying, "So only *one* of the boxes on the green side has a marble underneath." The side with more marbles (blue vs. green) was counterbalanced across participants.

Next, a cartoon image of a child appeared in the middle of the screen, and the experimenter introduced the agent, saying, "Now, this is my friend. I want to find out if my friend *knows* what's under all of the boxes. Hmm. To figure out if my friend *really knows* what's under all of the boxes, let's ask her to show us a box that has a marble underneath. And we can see if she gets it right. We can ask our friend to show us a marble on the blue side, or we can ask her to show us a marble on the green side." The experimenter continued on to the test question, saying, "I need your help! I need to find out if my friend *knows* what's under all of the boxes. Should I ask her to find a marble on the blue side, or on the green side?" After participants responded, the experimenter asked them to explain their choice.

The experimenter then asked participants, "And which one is *harder*? Is it harder to find a marble on the blue side, or on the green side?" The experimenter again asked participants to explain their response, and finally asked the pre-registered inclusion questions, saying, "And can you remind me: which side had a lot of marbles? Blue or green? And which side only

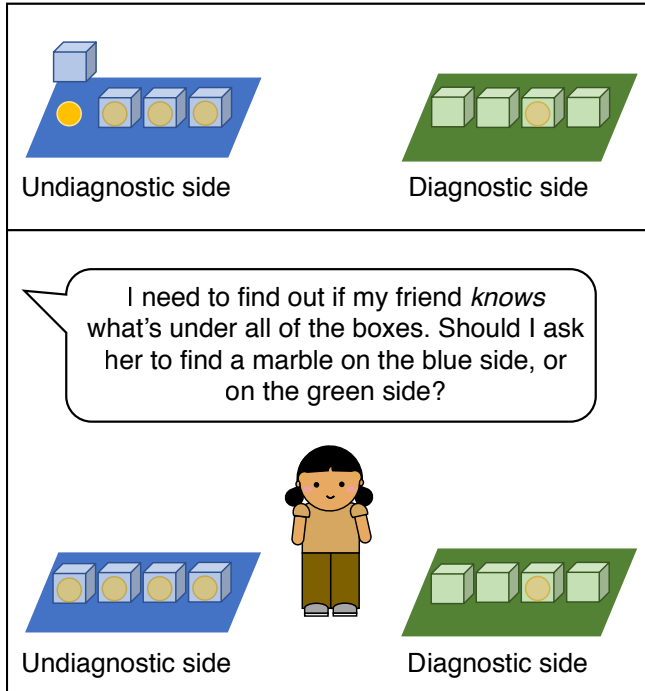


Figure 1. Experimental procedure. Each box was lifted one at a time to reveal its contents. Once participants saw the contents of all the boxes, the experimenter introduced a new friend and asked the test questions. Note that while we show each box's contents for clarity, in the experiment all boxes were opaque and participants could not see inside.

had one marble? Blue or green?" Note that although the blue side was always referenced first throughout the experiment, we counterbalanced whether this side was the correct option.

Results

For the 87.1% of participants whose sessions were video or audio taped ($n = 88/101$), two coders who were not involved in data collection determined exclusions according to pre-registered criteria. The first coder, blind to participant answers, determined whether the experiment was run correctly. The second coder, blind to condition, coded participant answers. The experimenter took notes on any deviations from the procedure, and for participants who were not video or audio taped the first author determined exclusions by comparing these notes to the pre-registered inclusion criteria. Eleven participants were recruited but not included in the final sample due to experimenter error ($n=3$), technical difficulties ($n=2$), because the participant did not provide codable answers to one or more questions ($n=2$), failed an inclusion question ($n=1$), was distracted ($n=1$), did not wish to continue ($n=1$), or due to interference ($n=1$).

Out of the final 90 participants included in the study, only 57.8% of participants chose to evaluate the agent's knowledge by asking her to find a marble on the more diagnostic side (where only one of the four boxes had a

marble underneath). This proportion is not reliably higher than chance ($n = 52$ of 90; 95% CI: 47.7 – 67.8). However, a logistic regression predicting performance as a function of age revealed a significant age difference ($\beta = 0.79$, $p = .003$), and performance within each age group qualitatively differed. Only 36.7% of four-year-olds ($n = 11$ of 30; 95% CI: 20 – 53.3) and 56.7% of five-year-olds ($n = 17$ of 30; 95% CI: 40 – 73.3) preferred to ask about the diagnostic side, whereas 80% of six-year-olds ($n = 24$ of 30; 95% CI: 66.7 – 96.7) did so (see Figure 2).

While only six-year-olds reliably wanted to ask the agent about the more diagnostic side, children of all ages understood that it was harder to find a marble on this side. 90% of participants ($n = 81$ of 90) correctly identified that it would be harder to find a marble on the diagnostic side, a proportion reliably higher than chance (95% CI: 84.1 – 96.6). A logistic regression predicting performance as a function of age did not reveal any significant age difference ($\beta = 0.35$, $p = 0.39$), and performance within each age group was qualitatively similar. 83.3% of four-year-olds ($n = 25$ of 30; 95% CI: 70 – 96.7), 96.7% of five-year-olds ($n = 29$ of 30; 95% CI: 93.3 – 100), and 90% of six-year-olds ($n = 27$ of 30; 95% CI: 80 – 100) judged that it would be harder to find a marble on the diagnostic side (see Figure 2).

Finally, participants' explanations were coded post-hoc (not pre-registered) by the first author and another experimenter. Coders identified whether participants explicitly compared the diagnostic and undiagnostic sides, justifying their chosen side in reference to the other (e.g., "because it has more of the marbles"; "because it is more tricky"; "because there is only one marble under there") — or whether participants' answers simply described their chosen side, without explicit reference to the other (e.g., "because I saw four marbles under it"; "because it's easy"; "because there is a marble under there"). Uncodable explanations were designated as "other". Inter-rater reliability was high for both explanation types (test question: 86.7%; Cohen's $\kappa = 0.8$; $p < .001$; "what's harder" question: 97.8%; Cohen's $\kappa = 0.97$; $p < .001$). Disagreements were resolved by discussion.

Six-year-olds often compared the two sides, both when explaining which side they wanted to ask about ($n = 20$ of 30), and when explaining where it was harder to find a marble ($n = 15$ of 30). Unsurprisingly, almost every six-year-old who justified their response by referencing both sides had answered the relevant test question correctly (see Figure 2).

In contrast, the relation between younger children's explanations and their responses qualitatively differed between the two test questions. Four- and five-year-olds who referenced both sides when explaining which was harder ($n = 21$ of 60) had all correctly identified the diagnostic side as more difficult. But most four- and five-year-olds who reference both sides when explaining which one they wanted to ask the agent about ($n = 15$ of 60) had actually asked about the incorrect undiagnostic side (9 of the 15; see Figure 2).

This suggests that some younger children understood that both sides were not equally informative to ask about, but did not reliably consider or use this information when deciding

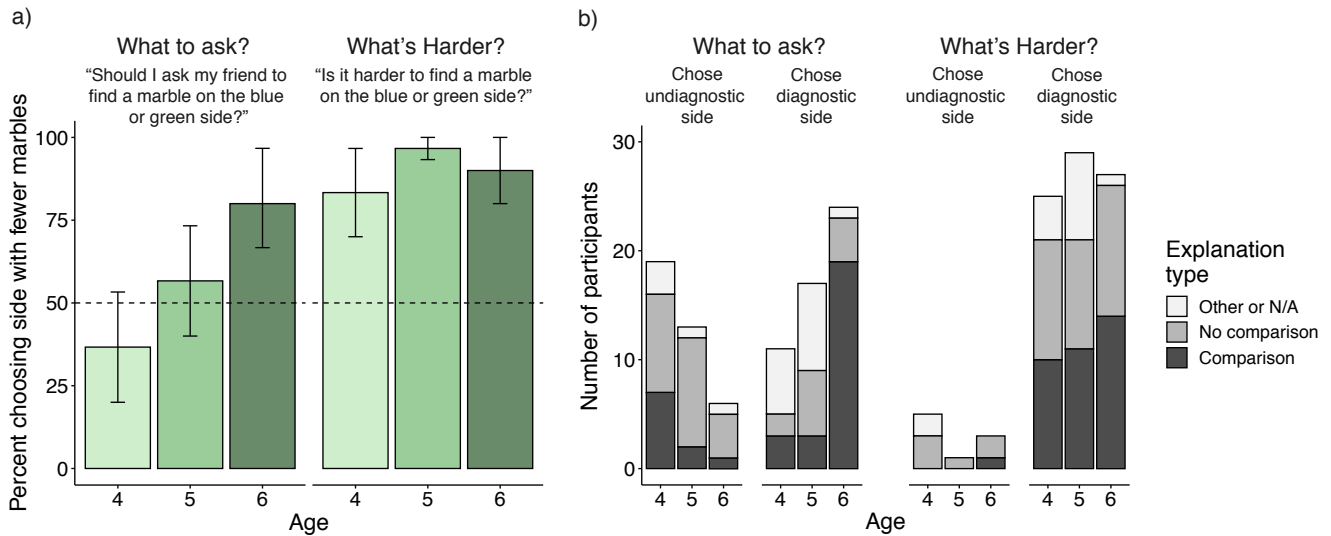


Figure 2. (a) Participant choices visualized by age group. The dotted line indicates predicted chance performance. Vertical bars show 95% bootstrapped confidence intervals. (b) Distribution of explanation types by test question and age group. “No comparison” explanations referred only to the chosen side; “comparison” explanations directly or indirectly referenced both sides.

how to evaluate the agent’s knowledge state. However, children seemed quite able to use the same information to judge which side was harder, suggesting that young children may struggle to use probability to make epistemic judgments, but not objective judgments about features like difficulty.

General Discussion

The capacity to teach or learn, help or hinder, and even punish or forgive, relies at least in part on an understanding of what others know. But the link between knowledge and action is highly variable, and is often mediated by the state of the world. For instance, if a friend arrives at their subway stop just as the train pulls into the station, you might infer they knew exactly when the train was due (and timed their walk accordingly). If they only caught the train because it was running late, you might be less charitable in your epistemic attribution. And if you find out that this train is actually late most of the time, then you might again suspect that your friend knows exactly what they’re doing.

Successful knowledge inferences thus require us to consider others’ actions in context, comparing how different degrees of knowledge might lead them to act in any given situation. We show that by the end of preschool, children understand how epistemic states and the environment interact, deciding what others know by comparing the likelihood of their actions under different degrees of knowledge. Specifically, by age six, children are sensitive to an agent’s expected probability of success under different epistemic states, understanding that a task where random success is improbable will better reveal what an agent knows.

Prior research shows that children are sensitive to probability from infancy (Denison et al., 2013; Xu & Garcia, 2008; Gweon et al., 2010), and infer others’ intentions and desires by considering the probability of their action outcomes from the first years of life (Diesendruck et al., 2015; Kushnir et al., 2012). In contrast, however, we find that children consider the probability of an action when evaluating epistemic states only by age six. And these results are consistent with related work suggesting that the ability to integrate probability and belief to predict others’ emotions is still developing between age five and seven (Doan, et al., 2018; MacLaren & Olson, 1993; Ruffman & Keenan, 1996; but see Scott, 2017). For instance, six-year-olds can accurately predict whether an agent will be surprised when asked to reason about the objective probability of an outcome—but not when prompted to consider an agent’s belief over this outcome (Doan et al., 2018). Taken together, these results suggest that even young children can judge whether an outcome is likely, but struggle to consider probability when predicting how others’ epistemic states will lead them to act or react.

This apparent divide in children’s use of probability in mental-state reasoning is consistent with the broader development of children’s Theory of Mind. While children can infer others’ goals, intentions, and desires from the first years of life (Gergely & Csibra, 2003; Meltzoff, 1995; Woodward, 1998), it is not until age four or five that children can reliably and explicitly represent others’ false beliefs (Wellman et al., 2001). Younger preschoolers may thus have less experience reasoning about epistemic states, as compared to mental states like desires, and it may take them longer to fully understand how epistemic states and action

relate. This possibility is consistent with research finding that a full understanding of epistemic states may continue to develop even past the preschool years: for instance, it is not until age six that children even begin to appreciate that ignorant agents will search randomly between locations (Ruffman 1996; Chen et al., 2015; Friedman & Petrashek, 2009). This possibility highlights the need for further research to investigate not only how children represent beliefs, but also how young children reason about and infer others' epistemic states (see also Phillips et al., 2020).

Our work also leaves open at least three further questions. First, four- and five-year-olds showed no reliable preference to ask the agent to complete the more diagnostic task. To better understand why this might be, it is useful to consider the components necessary for success. Participants who rely on Theory of Mind to solve our task must first notice that one side has more marbles than the other. Second, they need to consider the probability of finding a marble on each side given different knowledge states (or at the very least, whether the agent is more likely to find a marble on one side given ignorance). Finally, children need to use this information to decide what will be most diagnostic to ask.

Our participants likely did not stumble at the first step: prior research suggests that by five or six months of age, even infants distinguish between sets that differ in magnitude by 50% or more (Wynn et al., 2002; Xu & Spelke, 2000). Consistent with this, 90% of younger children explicitly judged that the diagnostic task was harder, showing clear evidence that they distinguished between the two. Instead, younger children may have struggled to consider how ignorance could affect an agent's probability of success (especially as the prior work reviewed above suggests that four- and five-year-olds may struggle to reason about ignorance). Or, younger children may have understood that an ignorant agent's probability of success differed amongst the two tasks, but failed to use this information when deciding what would be most diagnostic to ask. Indeed, neither task is fully diagnostic; an ignorant agent could succeed by chance even on the diagnostic task. So younger children may have struggled to compare these two options and select the more informative of the two.

However, it is also possible that children struggled because they did not want to ask the agent to complete a difficult task. Children begin acting on prosocial motivations early in life (Warneken & Tomasello, 2006); younger children might have considered the agent's probability of success on each task and then decided to help her (rather than trying to select the task that would best reveal her knowledge state). Similarly, if younger children misunderstood the test question or attended only to the last thing they heard ("Should I ask her to find a marble on the blue side, or on the green side?"), a motivation to assist the agent could cause children to ask her about the undiagnostic side (where success is assured).

To address this possibility, in an ongoing experiment we test whether four- to six-year-olds judge that an improbable success is more likely to indicate prior knowledge than an

assured success. Specifically, participants observe one agent find a marble on the more diagnostic side (where only one box has a marble) and another agent find a marble on the undiagnostic side (where every box has a marble underneath). Participants are told that one agent already looked under the boxes on her side, and are asked to judge who looked under the boxes. Here, participants cannot help either agent, and they are explicitly asked to make an epistemic judgment. If we observe a similar developmental trajectory in this experiment, this would suggest that younger children's performance cannot be explained entirely by a desire to be helpful. However, if younger children succeed, this would suggest that their difficulties may have been task-specific.

Second, we found that by age six, children understand that a task with a low probability of random success will better reveal an agent's knowledge state (as compared to a task where success is assured). However, it is possible that even older children's reasoning was not fully probabilistic—participants could have simply chosen the only side where failure was possible (without considering precisely how probable it was). Or participants could have relied on an even simpler heuristic, assuming that ignorant agents make mistakes (e.g., Ruffman, 1996), and thus selecting the only side where error was possible.

Although six-year-olds' explanations suggest that many participants did explicitly compare the agent's probability of success on each task, ongoing work directly tests for this possibility. Specifically, we are running a version of the current experiment, with the difference that on the "undiagnostic" side, only three of the four boxes have marbles underneath. We test whether children still prefer to ask the agent to find a marble on the more diagnostic side.

Finally, it is worth noting that even as adults, we do not always infer knowledge from an otherwise unlikely success (most of us would agree that a person who picked the winning lottery number didn't know they would win). Conversely, even adults can struggle to correctly estimate the probability of a successful action under ignorance (for instance, attributing otherworldly knowledge when a vague astrological prediction or palm reading happens to describe their life). And sometimes we rely on trait attributions (rather than epistemic or world states) to explain others' successes or failures; for instance, describing a person as "unlucky" if they have experienced a string of otherwise unexpectedly poor outcomes. It is unclear how young children reason about such edge cases, especially in situations where even adults can struggle.

Conclusion

To infer the cause of a failed action, figure out what to teach, or even decide who knew better, we must understand others' epistemic states. In the current work, we find that by age six, children understand that the state of the world mediates the relation between knowledge and action—using this to decide under what conditions an action or outcome truly reveals knowledge. These results highlight the complexity of everyday epistemic judgments, and the need

for further research into children's understanding of the relation between knowledge, ignorance, belief, and action.

Acknowledgments

We thank Colin Jacobs, Julianna Lu, Sofia Rubio, Eden Senay, and Hudson Patterson for assistance with data collection. We thank Ilayda Orhan and Rodney Tompkins for assistance with coding. We thank Katie Vasquez for her collaborative approach to participant recruitment. This work was supported by Yale's Franke Program in Science and the Humanities, via a Franke Interdisciplinary Graduate Award to RA.

References

- Aboody, R., Flowers, M., Zhou, C., & Jara-Ettinger, J. (2019). Ignorance = doing what is reasonable: Children expect ignorant agents to act based on prior knowledge. *Proceedings of the 41st Annual Conference of the Cognitive Science Society* (pp. 1297-1303). Montreal, QB: Cognitive Science Society.
- Aboody, R., Huey, H., & Jara-Ettinger, J. (2018). Success does not imply knowledge: Preschoolers believe that accurate predictions reveal prior knowledge, but accurate observations do not. In Kalish, C., Rau, M., Zhu, J., & Rogers, T. (Eds.), *Proceedings of the 40th Annual Conference of the Cognitive Science Society* (pp. 76-81). Austin, TX: Cognitive Science Society.
- Aboody, R., Zhou, C., & Jara-Ettinger, J. (2021). In pursuit of knowledge: Preschoolers expect agents to weigh information gain and information's cost when deciding whether to explore. *Child Development*.
- Bridgers, S., Buchsbaum, D., Seiver, E., Griffiths, T. L., & Gopnik, A. (2016). Children's causal inferences from conflicting testimony and observations. *Developmental Psychology*, 52(1), 9.
- Chen, Y., Su, Y., & Wang, Y. (2015). Young children use the "ignorance=getting it wrong" rule when predicting behavior. *Cognitive Development*, 35, 79-91.
- Corriveau, K., & Harris, P. L. (2009). Preschoolers continue to trust a more accurate informant 1 week after exposure to accuracy information. *Developmental Science*, 12(1), 188-193.
- Denison, S., Reed, C., & Xu, F. (2013). The emergence of probabilistic reasoning in very young infants: Evidence from 4.5- and 6-month-olds. *Developmental Psychology*, 49(2), 243.
- Diesendruck, G., Salzer, S., Kushnir, T., & Xu, F. (2015). When choices are not personal: The effect of statistical and social cues on children's inferences about the scope of preferences. *Journal of Cognition and Development*, 16(2), 370-380.
- Doan, T., Friedman, O., & Denison, S. (2018). Beyond belief: The probability-based notion of surprise in children. *Emotion*, 18(8), 1163.
- Einav, S., & Robinson, E. J. (2011). When being right is not enough: four-year-olds distinguish knowledgeable informants from merely accurate informants. *Psychological Science*, 22, 1250-1253.
- Friedman, O., & Petrashek, A. R. (2009). Children do not follow the rule "ignorance means getting it wrong". *Journal of Experimental Child Psychology*, 102(1), 114-121.
- Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naïve theory of rational action. *Trends in Cognitive Sciences*, 7, 287-292.
- Gweon, H., Tenenbaum, J. B., & Schulz, L. E. (2010). Infants consider both the sample and the sampling process in inductive generalization. *Proceedings of the National Academy of Sciences*, 107(20), 9066-9071.
- Harris, Paul L. *Trusting what you're told: How children learn from others*. Harvard University Press, 2012.
- Jara-Ettinger, J., Floyd, S., Tenenbaum, J. B., & Schulz, L. E. (2017). Children understand that agents maximize expected utilities. *Journal of Experimental Psychology: General*, 146, 1574.
- Koenig, M. A., & Harris, P. L. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child Development*, 76(6), 1261-1277.
- Koenig, M. A., Clément, F., & Harris, P. L. (2004). Trust in testimony: children's use of true and false statements. *Psychological Science*, 15, 694-698.
- Kominsky, J. F., Langthorne, P., & Keil, F. C. (2016). The better part of not knowing: Virtuous ignorance. *Developmental Psychology*, 52(1), 31.
- Kushnir, T., Xu, F., & Wellman, H. M. (2010). Young children use statistical sampling to infer the preferences of other people. *Psychological Science*, 21(8), 1134-1140.
- 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, 1038-1041.
- MacLaren, R., & Olson, D. (1993). Trick or treat: Children's understanding of surprise. *Cognitive Development*, 8(1), 27-46.
- Meltzoff, A. N. (1995). Understanding the intentions of others: re-enactment of intended acts by 18-month-old children. *Developmental Psychology*, 31(5), 838.
- Nurmsoo, E., & Robinson, E. J. (2009b). Identifying unreliable informants: Do children excuse past inaccuracy? *Developmental Science*, 12, 41-47.
- Phillips, J., Buckwalter, W., Cushman, F., Friedman, O., Martin, A., Turri, J., ... & Knobe, J. (2020). Knowledge before belief. *Behavioral and Brain Sciences*, 1-37.
- Ruffman, T., & Keenan, T. R. (1996). The belief-based emotion of surprise: The case for a lag in understanding relative to false belief. *Developmental Psychology*, 32(1), 40.
- Ruffman, T. (1996). Do children understand the mind by means of simulation or a theory? Evidence from their understanding of inference. *Mind & Language*.
- Saxe, R. (2005). Against simulation: the argument from error. *Trends in Cognitive Sciences*, 9(4), 174-179.

- Scott, R. M. (2017). Surprise! 20-month-old infants understand the emotional consequences of false beliefs. *Cognition*, 159, 33-47.
- Warneken, F., & Tomasello, M. (2006). Altruistic helping in human infants and young chimpanzees. *Science*, 311(5765), 1301-1303.
- Wellman, H. M. (2014). *Making minds: How theory of mind develops*. Oxford University Press.
- Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, 72(3), 655-684.
- Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, 69(1), 1-34.
- Wu, Y., & Schulz, L. E. (2018). Inferring beliefs and desires from emotional reactions to anticipated and observed events. *Child Development*, 89(2), 649-662.
- Wynn, K., Bloom, P., & Chiang, W. C. (2002). Enumeration of collective entities by 5-month-old infants. *Cognition*, 83(3), B55-B62.
- Xu, F., & Garcia, V. (2008). Intuitive statistics by 8-month-old infants. *Proceedings of the National Academy of Sciences*, 105, 5012-5015.
- Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition*, 74(1), B1-B11.