Child Development

CHILD DEVELOPMENT

EMPIRICAL ARTICLE

Six-Year-Olds, but Not Younger Children, Consider the Probability of Being Right by Chance When Inferring Others' Knowledge

Rosie Aboody[1] | Julianna Lu[2] | Stephanie Denison[3] | Julian Jara-Ettinger[1]

[1]Yale University, New Haven, Connecticut, USA | [2]University of Toronto, Toronto, Ontario, Canada | [3]University of Waterloo, Waterloo, Ontario, Canada

Correspondence: Rosie Aboody (rosie.aboody@yale.edu)

Received: 8 August 2024 | **Revised:** 2 March 2025 | **Accepted:** 19 April 2025

Funding: This work was supported by NSF award BCS-2045778 to JJE, and a Franke Interdisciplinary Graduate Award from Yale's Franke Program in Science and the Humanities to RA.

Keywords: epistemic reasoning | knowledge | probability | social cognition | Theory of Mind

ABSTRACT

When determining what others know, we intuitively consider not only whether they succeed but also their probability of success in the absence of knowledge (e.g., random guessing). Across three experiments ($n = 240$ North American 4–6-year-olds, data collected between 2020–2023) we find that 4-year-olds understand that tasks with a lower probability of chance success are harder. However, it is not until age 6 that children use this understanding to gauge (Experiment 1) and infer (Experiments 2–3) what others know. These results suggest that, although basic probabilistic reasoning and representations of knowledge are well in place by age 4, children do not integrate the two to make mental-state inferences until much later, pointing to an area of important developmental change in Theory of Mind.

1 | Introduction

Many of our social interactions are shaped by what we think other people know. Representing other people's knowledge helps us predict their behavior (e.g., Wellman et al. 1990; Wellman 2014; see also Jara-Ettinger et al. 2017), work cooperatively by sharing and requesting relevant information (O'Neill 1996; Bridgers et al. 2020), and even determine others' moral culpability when their actions cause harm (e.g., Young et al. 2007). However, building accurate representations of what others know can be difficult because people do not have access to each other's minds and must instead infer knowledge from observable action.

This problem might be particularly critical in early childhood. Given that children heavily rely on others to learn about the world, much work has argued that they must have mechanisms for distinguishing those who are knowledgeable from those who are not (Harris et al. 2018; Sperber et al. 2010). Yet, while the capacity to represent other people's knowledge emerges in infancy (Onishi and Baillargeon 2005; Ting et al. 2021), this capacity is brittle and preschoolers often make counter-intuitive inferences about knowledge (Aboody et al. 2019; Brosseau-Liard and Birch 2010; Crivello and Poulin-Dubois 2018; Chuey, Sparks, et al. 2023; Kampis et al. 2021; Schuwerk et al. 2018). For instance, young preschoolers do not believe that an agent who takes a nonobvious action to activate a toy is probably more knowledgeable than someone who succeeds by taking the obvious action (Aboody et al. 2019). Children around this age also overextend their inferences, assuming that more knowledgeable people must also be nicer (Brosseau-Liard and Birch 2010).

A large body of work suggests that children's knowledge inferences, like the ones we reviewed above, reveal that they rely on a simple cue to knowledge: accuracy. Thus, a challenging inferential problem—determining what others know based on indirect observable behavior—becomes an easy one by simply assuming that accurate people are knowledgeable and inaccurate people are not (Corriveau and Harris 2009; Koenig et al. 2004; Koenig and Harris 2005; Harris 2012). For instance, 4-year-olds reliably distinguish accurate from inaccurate agents, but they do not always distinguish between degrees of inaccuracy, treating someone who gets one of four questions wrong as equally untrustworthy as someone who gets three out of four questions wrong (Pasquini et al. 2007). Similarly, preschoolers often fail to appreciate that not all errors are equal. For example, when observing one blindfolded and one nonblindfolded agent both mislabel objects, preschoolers treat these agents as equally poor sources of information, failing to excuse the poor performance of the blindfolded agent (Nurmsoo and Robinson 2009; see also Bridgers et al. 2016).

While treating accuracy as a cue to knowledge (or inaccuracy as a cue to ignorance) is a helpful heuristic, the two are not always in perfect correspondence. As adults, we are intuitively sensitive to the probability that someone would be accurate under different knowledge states. For example, suppose you're talking to a colleague who is trying to convince you that he met your sister when he was in college. If your sister studied in a different country, you might reasonably be skeptical. To convince you, your colleague should show that they can produce accurate statements they would be unlikely to get right by chance, like unique knowledge of her tattoo, or unusual hobbies, and not generic information anyone could have guessed, like that she shares your eye color or speaks your native language. That is, considering the relative likelihood of being accurate by chance can help us infer what other people know.

The limits of the knowledge-accuracy heuristic pose an important theoretical problem. If children circumvent complex reasoning about beliefs through this heuristic, then their capacity to track knowledgeable agents is limited and only approximately correct, often leading to incorrect representations. This could suggest that children's need for epistemic vigilance—having accurate representations of who is knowledgeable and who is not—might not be as critical (potentially leading this capacity to be impoverished), as children successfully get by in their early social learning using only a coarse sense of what others know.

At the same time, there are three reasons to believe that young preschoolers might be able to move beyond a simple knowledge-accuracy heuristic and consider the possibility that an agent would be accurate by chance. First, studies that explicitly control for accuracy have revealed that children's understanding is more nuanced than what the simple heuristic predicts. For instance, children distinguish between an agent who is accurate because they received help and an agent that is independently accurate (Einav and Robinson 2011). Similarly, children distinguish between agents that make accurate predictions and agents that merely make accurate observations, and they causally attribute prior experiences and additional knowledge to the agent making accurate predictions (Aboody et al. 2022). Thus, these studies suggest that young children already have a basic causal understanding of how knowledge relates to action, which could extend to include considerations of how likely an agent is to be accurate under different epistemic states.

A second reason why children might be able to consider the likelihood of being accurate under different knowledge states is because this is a form of probabilistic reasoning, which even infants are sensitive to (Denison et al. 2013; Xu and Garcia 2008; Gweon et al. 2010). Finally, a third reason is that related research has already found that young children can integrate probabilistic reasoning when making inferences about others' desires, inferring that rarer choices imply stronger preferences (Diesendruck et al. 2015; Kushnir et al. 2010; although, interestingly, the ability to do the same with emotion inferences develops relatively late: Doan et al. 2023).

Together, these arguments, combined with a commonly held belief that epistemic vigilance is critical in childhood, would suggest that children should be able to consider how easy or difficult it is to be accurate by chance and use this to infer how much someone knows. Thus, despite a strong literature suggesting that children make epistemic judgments by relying on a coarse and insensitive heuristic where accuracy inevitably implies knowledge, young children might be sensitive to the probabilistic relation between accuracy and knowledge, enabling them to only attribute knowledge when an agent's accuracy warrants it. In the current work, we investigate this question, testing whether young children consider the probability of being accurate by chance when inferring knowledge or whether they treat accuracy as a context-insensitive cue to knowledge.

In Experiment 1, 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. In Experiment 2, we test whether children are more likely to attribute knowledge to an agent who successfully completes the same "diagnostic" task (compared to the "undiagnostic" task). And in Experiment 3, we test whether children's reasoning is truly probabilistic, replicating the method of Experiment 2 but comparing the same task with a 25% chance of random success to one where random success is merely probable (75%), rather than assured.

We focus on 4–6-year-olds for two reasons. First, this is a period where children are becoming more active members of the social world. They are attending preschool and kindergarten where they have increasing access to social learning opportunities from a more varied collection of people, including their parents, siblings, classmates, and teachers—and some of these sources can be unreliable. Second, this is an age where children's belief reasoning (Wellman et al. 2001; Wellman 2014; Wu and Schulz 2018) and understanding of ignorance (Chen et al. 2015; Fabricius et al. 2021; Friedman and Petrashek 2009; German and Leslie 2001; Ruffman 1996; Saxe 2005) is still developing and thus it may be an age range where children begin to integrate contextual features, like probabilistic reasoning, into their reasoning about knowledge.

2 | Approach to Analyses and General Methods

Consistent with recent recommendations for statistical best practices, we take an estimation approach to data analysis (Cohen 1994; Cumming 2014). We estimate effect sizes by bootstrapping our data and obtaining 95% confidence intervals, taking confidence intervals that do not cross chance as evidence of a reliable effect.

The sample sizes, procedures, predictions, exclusion criteria, and analysis plan for all experiments were preregistered. All preregistrations, stimuli, data, and analysis files are available in the OSF project page: https://osf.io/fm8e2. All experiments were IRB-approved. Note that while our analysis plan was not exploratory—all analyses were preregistered—our research questions can be considered exploratory because we did not have exact predictions about the precise developmental trajectory or age at which children would begin to succeed.

The sample size for Experiment 1 was determined through a Monte Carlo power analysis prior to preregistration, and we maintained this sample size in Experiments 2–3. Assuming 75% of participants are attentive (and thus answer the test question correctly), a sample of 30 participants per age group yields power = 0.899.

3 | Experiment 1

3.1 | Method

3.1.1 | Participants

Ninety 4–6-year-olds (mean age: 5.51 years, range: 3.97–6.9 years) were recruited and tested online, via a video-chat research platform. Eleven additional participants were recruited but not included in the study (see Results). Note that, because we were uncertain what developmental trajectory we would find, we initially preregistered running 4- and 5-year-olds; we preregistered our 6-year-old sample about 10 weeks later. All preregistrations can be found in the project OSF page.

There is little evidence that basic Theory of Mind and probabilistic reasoning capacities vary across race/ethnicity or gender; therefore, we did not preregister collection of these participant demographics. However, research conducted online often increases sample diversity by broadening participation beyond the immediate vicinity of elite institutions (Sheskin et al. 2020). All data were collected between July 2020 and February 2021.

3.1.2 | 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).

3.1.3 | Procedure

Figure 1 shows the experimental procedure. We counterbalanced which side had one marble and which side had three marbles, but all other aspects of the procedure and stimuli were fixed.

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 and replaced 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

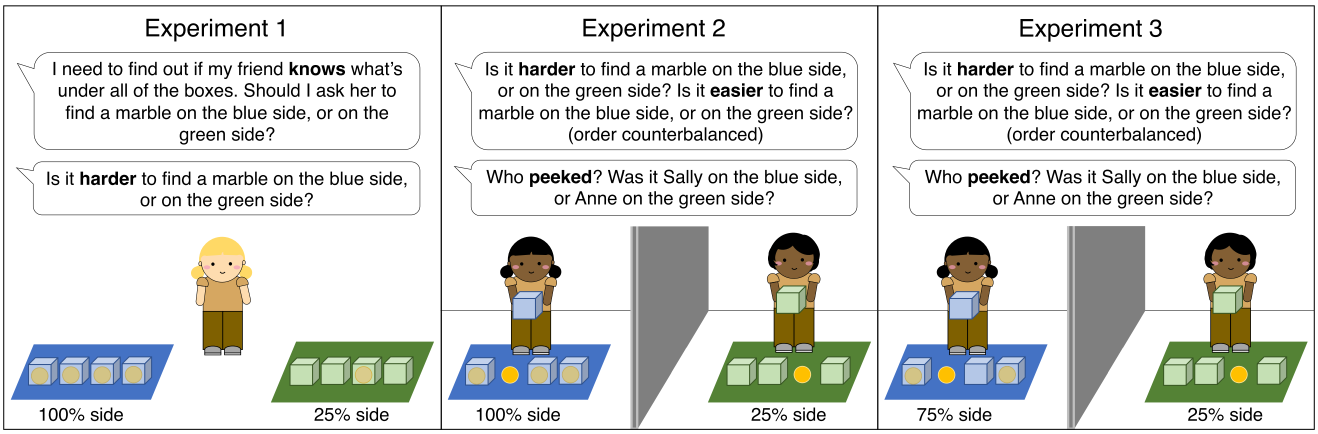

FIGURE 1 | Procedure for all experiments. In all experiments, participants were first shown the contents of every box (not pictured); we counterbalanced which side had more marbles (blue/green). Note that all boxes are transparent here for visualization purposes but were opaque during the actual experiments. In Experiment 1, the experimenter introduced a new friend and asked participants to gauge her knowledge by asking her to find a marble on either the blue or green side. Participants were then asked to judge whether it was harder to find a marble on the blue side or the green side. In Experiment 2, the experimenter introduced two new friends and explained that one had already peeked under the boxes on her side, and that one had not looked under the boxes. The experimenter asked each in turn to find a marble on her side; both immediately succeeded. Here, participants were first asked what was harder and what was easier (counterbalanced), and next were asked to identify which one of the two friends had peeked and already knew what was under the boxes. Finally, Experiment 3 replicated Experiment 2 in a fully probabilistic case. We removed one marble from the undiagnostic side, yielding a "less diagnostic" side with 75% odds of random success.

under this box!" After lifting all 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 on to 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), whereas 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 is under all the boxes. Hmm. To figure out if my friend *really knows* what is under all 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 to the test questions, 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 preregistered inclusion questions, saying, "And can you remind me: which side had a lot of marbles? Blue or green? And which side only had one marble? Blue or green?"

3.2 | Results

For the 87.1% of participants whose sessions were videotaped ($n=88/101$), two coders who were not involved in data collection determined exclusions according to preregistered 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 preregistered 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 the inclusion question ($n=1$), was distracted ($n=1$), did not wish to continue ($n=1$), or due to family interference ($n=1$).

Out of the final 90 participants included in the study, only 57.8% chose to ask about the 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: 47.8–67.8). However, a logistic regression predicting performance as a function of centered age (continuous) revealed a significant effect of age ($\beta=0.79$, $p=0.003$), and performance within each age group qualitatively differed. Only 36.7% of 4-year-olds ($n=11$ of 30; 95% CI: 20–53.3) and 56.7% of 5-year-olds ($n=17$ of 30; 95% CI: 40–73.3) preferred to ask about the diagnostic side, whereas 80% of 6-year-olds ($n=24$ of 30; 95% CI: 66.7–96.7) did so (see Figure 2).

While only 6-year-olds reliably chose the 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: 84.4–96.7). A logistic regression predicting performance

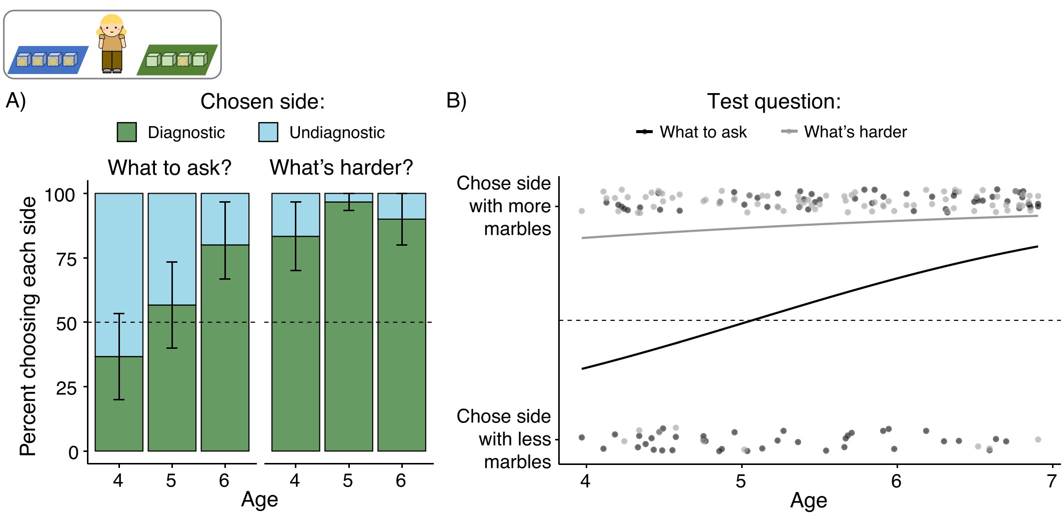

FIGURE 2 | At top left, see a basic procedure schematic. (A) Participant choices visualized by age group. The dotted line indicates predicted chance performance. Vertical bars show 95% bootstrapped confidence intervals. (B) Participant choices plotted continuously by age, along with a logistic regression fitted to each data set. Points are jittered along the *Y*-axis (but not the *X*-axis).

as a function of centered age (continuous) did not reveal any significant effect of age ($\beta = 0.35$, $p = 0.39$), and performance within each age group was qualitatively similar. 83.3% of 4-year-olds ($n = 25$ of 30; 95% CI: 70–96.7), 96.7% of 5-year-olds ($n = 29$ of 30; 95% CI: 93.3–100), and 90% of 6-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). See Supporting Information for participants' raw explanations as well as an analysis of these explanations. Finally, a post hoc power analysis with an 80% pass rate (based upon 6-year-olds' performance on our test question, deciding which side to ask about) yields power = 0.97.

3.3 | Discussion

The results from Experiment 1 suggest that 4- to 6-year-olds understand that it would be harder to find a marble on the diagnostic side (probability = ¼) relative to the undiagnostic side (probability = 1), but only 6-year-olds reliably selected the diagnostic side to assess the agent's knowledge. These results hint that even young children may be able to use probability to make objective judgments about features like difficulty—but not to gauge others' epistemic states.

There are at least three possible alternative explanations behind younger children's failures. A first possibility is that younger children did not initially consider the relative difficulty of the two sides (possibly due to task novelty, other task demands, or a failure to appreciate the relation between difficulty and success *spontaneously*), but would have successfully integrated this information if they had been asked to consider it before the test question.

A second possibility is that the breakdown in younger children's performance was specifically tied to making predictions. Perhaps asking children to infer knowledge based on an action the agent takes could be easier. Indeed, some research suggests that, at least in certain cases, young children can identify what mental states caused a behavior earlier than they can predict how the same mental states might lead an agent to act (see Wellman 2011). So, younger children might have struggled not with integrating epistemic states and probability, but simply with being asked to make an action prediction.

Finally, it is also possible that younger children understood precisely how different epistemic states might affect an agent's probability of success on each side but did not want to ask the agent to complete a difficult task. Children begin acting on prosocial motivations early in life (Warneken and Tomasello 2006); younger children might have decided to help the character in our experiment (rather than trying to select the task that would best reveal her knowledge state), though admittedly it is not clear why younger and not older children might have taken this approach. Experiment 2 addresses these possibilities.

4 | Experiment 2

Experiment 2 was a conceptual replication of Experiment 1, modified to address potential alternative explanations. Here, children saw an agent find a marble on an undiagnostic side (where all four boxes had a marble underneath), and another agent find a marble on a diagnostic side (where only one of the four boxes had a marble underneath). Children were first asked to identify the side where it would be harder to find a marble, the side where it would be easier, and then to infer which of the two agents already knew the contents of the boxes.

4.1 | Method

4.1.1 | Participants

Ninety 4–6-year-olds (mean age: 5.5 years, range: 4.05–6.95 years) were recruited and tested online, via the same video-chat platform as in Experiment 1. Ten additional participants were recruited but not included in the study (see Results). As before, we did not preregister collection of participant demographics; all data were collected between July 2021 and February 2022.

4.1.2 | Stimuli

Stimuli consisted of a PowerPoint presentation, featuring cartoon characters of two girls, four blue boxes lined up on a blue background, four green boxes lined up on a green background, and a wall separating the two. 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).

4.1.3 | Procedure

Figure 1 shows the experimental procedure. The procedure began nearly identically to that of Experiment 1, with the exception that after drawing participants' attention to the boxes, the experimenter also pointed out the wall in the middle of the screen, saying, “And look! There's a big wall in the middle. Do you see the wall? Great!” Without a wall separating the two sides, a knowledgeable agent could have chosen to position herself near the undiagnostic side simply because she knew it was easier to find a marble on this side—the addition of the wall minimized this concern.

After pointing out the wall, the experiment always began with eight boxes appearing on the screen, and participants were introduced to the contents of the boxes in the same way as in Experiment 1 (the experimenter lifted each box one at a time). The only difference was that at the end, after lifting all the boxes, the experimenter again repeated, “So, all of the boxes on the blue side have a marble underneath, and only one of the boxes on the green side has a marble underneath.” We added this repetition as a conservative measure to emphasize the difference between the two sides (after younger children in Experiment 1 failed to recognize that the diagnostic side was more informative to ask about). As before, the side with more marbles (blue vs. green) was counterbalanced across participants.

Next, the experimenter said, “Here I have two friends.” An image of a cartoon child appeared on the left side of the screen, and the experimenter explained, “This is my friend Sally, on the *blue* side” (text italicized to mark words emphasized by the experimenter). Next, another child appeared on the right side of the screen, and the experimenter explained, "this is my friend Anne, on the *green* side." Continuing on, the experimenter said, "Now, right before you came here today, *one* of these friends peeked under the boxes. The other friend did *not* peek, and has never looked under the boxes. So *one* of these friends knows what's under the boxes on her side. And the other friend has no idea what's under the boxes on her side. I don't know if Sally peeked under all of the *blue* boxes, *or* if Anne peeked under all of the *green* boxes. Only *one* of my friends peeked. Hmm. To figure out which one of my friends peeked, let's ask each one a question."

The experimenter continued, "First, I'll ask my friend Sally. Let's figure out if Sally peeked under all the boxes on the *blue side*. Sally, can you find a marble on the blue side?" After asking this question, the experimenter showed participants that Sally had lifted the second box from the left, revealing a marble (see Figure 1). The experimenter said, "And look! Sally bent down and lifted this box. And she was right. So Sally found one of the boxes on the blue side that has a marble underneath!" Next, Sally put the box back, and the experimenter said, "Next, I'll ask my friend Anne. Let's figure out if Anne peeked under all the boxes on the *green side*. Anne, can you find a marble on the green side?" After asking this question, the experimenter showed participants that Anne had lifted the third box from the left, revealing a marble. The experimenter said, in the same tone as they used for Sally, "And look! Anne bent down and lifted this box. And she was right. So Anne found the only box on the green side that has a marble underneath!" Next, Anne put the box back.

The experimenter then reviewed the procedure, saying, "Remember how I told you that only *one* of these friends peeked under the boxes on her side? And the other one did *not* peek, and did not look under the boxes on her side? So only *one* of our friends knows what's under the boxes. Well, my friend Sally on the blue side found *one* of the boxes that has a marble underneath. And my friend Anne on the green side found the *only* box that has a marble underneath."

Next, the experimenter proceeded to the test questions. In contrast to Experiment 1, participants were first asked to identify (1) where it was harder to find a marble and (2) where it was easier (order counterbalanced). These changes were intended to prompt participants to consider the odds of success on both the blue and green side, and to do so before being asked to make an epistemic inference (with the intention to provide participants the best possible chance of making the appropriate inference).

The experimenter asked, "Can you tell me, which one is *harder*? Is it harder to find a marble on the blue side, or on the green side?" After repeating the participant's answer and eliciting an explanation, the experimenter asked, "And can you tell me, which one is *easier*? Is it easier to find a marble on the blue side, or on the green side?" The experimenter repeated the participant's answer and elicited an explanation. The order of these two questions was counterbalanced. Finally, the experimenter asked children to make an epistemic inference, saying, "Now [participant name], I need your help! Can you tell me: who peeked? Was it Sally on the blue side, or Anne on the green side?" Again, the experimenter repeated the participant's answer and elicited an explanation. And as before, after the test questions, the experimenter asked the preregistered inclusion questions, saying, "And can you remind me: which side had a lot of marbles? Blue or green? And which side only had one marble? Blue or green?"

4.2 | Results

For the 93% of participants whose sessions were video- or audiotaped ($n=93/100$), two coders who were not involved in data collection determined exclusions according to preregistered criteria, as in Experiment 1. Ten participants were recruited but not included in the final sample due to participant distraction/inattention ($n=4$), experimenter error ($n=1$), family interference ($n=1$) because the participant did not provide codable answers to one or more test questions ($n=1$), failed inclusion ($n=1$), did not wish to continue ($n=1$), or had already participated in the past and was accidentally invited again to participate ($n=1$).

Out of the final 90 participants included in the study, 65.6% judged that the agent who succeeded on the diagnostic task (finding the only marble on her side) had peeked under the boxes. This proportion is reliably higher than chance ($n=59$ of 90; 95% CI: 56.7–75.6). A logistic regression predicting performance as a function of centered age (continuous) further revealed a significant age effect ($\beta=0.61$, $p=0.035$), and performance within each age group qualitatively differed. Only 53.3% of 4-year-olds ($n=16$ of 30; 95% CI: 36.7–70) and 63.3% of 5-year-olds ($n=19$ of 30; 95% CI: 46.7–80) inferred that the agent who completed the diagnostic task had prior knowledge, whereas 80% of 6-year-olds did so ($n=24$ of 30; 95% CI: 66.7–96.7; see Figure 3).

As in Experiment 1, children of all ages were able to identify which task was harder and which was easier. Of our participants, 83.3% ($n=75$ of 90) correctly identified that it would be harder to find a marble on the diagnostic side. And 83.3% also correctly identified it would be easier to find a marble on the undiagnostic side. These proportions are reliably higher than chance (95% CI: 75.6–91.1). A logistic regression predicting performance as a function of centered age (continuous) did not reveal any significant effect of age for either question (what's harder: $\beta=0.47$, $p=0.185$; what's easier: $\beta=-0.06$, $p=0.87$). Consistent with this, performance within each age group was qualitatively similar: when identifying what was harder, 73.3% of 4-year-olds ($n=22$ of 30; 95% CI: 60–90), 90% of 5-year-olds ($n=27$ of 30; 95% CI: 80–100), and 86.7% of 6-year-olds ($n=26$ of 30; 95% CI: 76.7–100) correctly judged that it would be harder to find a marble on the diagnostic side (see Figure 3). And when identifying what was easier, 83.3% of participants in each age group ($n=25$ of 30 each, 4-, 5-, and 6-year-olds; 95% CI: 70–96.7) judged that it would be easier to find a marble on the undiagnostic side (see Figure 3). See Supporting Information for explanations. Finally, a post hoc power analysis with an 80% pass rate (based upon 6-year-olds' performance on our test question, inferring who had peeked) yields power = 0.97.

4.3 | Discussion

This experiment suggests that by age 6 (but not before), children realize that an ignorant agent is unlikely to immediately

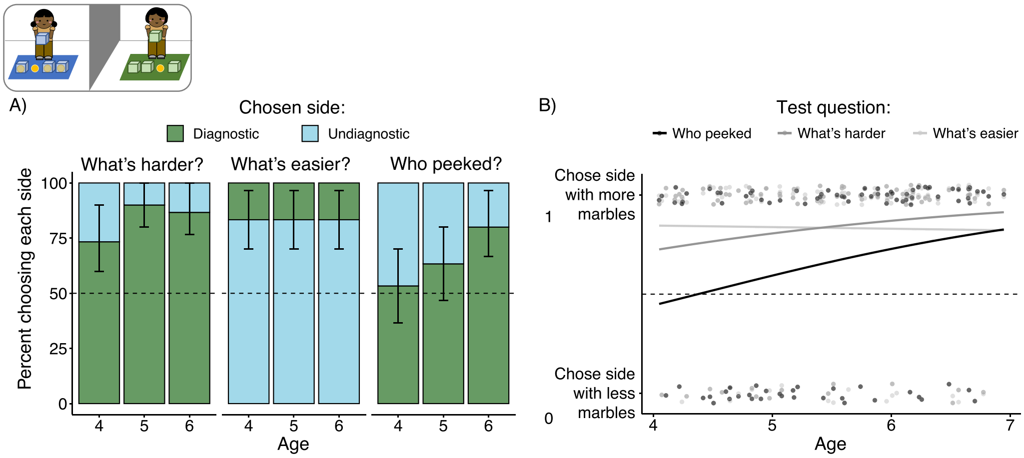

FIGURE 3 | At top left, see a basic procedure schematic. (A) Participant choices visualized by age group. The dotted line indicates predicted chance performance. Vertical bars show 95% bootstrapped confidence intervals. (B) Participant choices plotted continuously by age, along with a logistic regression fitted to each data set. Points are jittered along the *Y*-axis (but not the *X*-axis).

succeed on a task with a low probability of success (compared to a task where success is assured). Thus, 6-year-olds could infer prior knowledge from an otherwise improbable success. By contrast, 4- and 5-year-olds were able to identify that the diagnostic task would be harder to complete (and the undiagnostic task easier), but they were split when asked to infer which agent was knowledgeable.

These results demonstrate three things: First, children's difficulty integrating the likelihood of success into their epistemic reasoning is not specifically tied to making predictions (Experiment 1), and it extends to epistemic inferences as well (Experiment 2). Second, children's failure cannot be explained by appealing to a motivation to see agents succeed. While in Experiment 1, young children could have failed because they prioritized seeing the agent succeed, this account cannot explain their failure in the current experiment, as participants were not offered a chance to help either agent and were asked to infer who was knowledgeable after already seeing both agents succeed. Finally, children's general failure is not due to a specific failure to compare the relative probability of success. In the current experiment, participants were asked to judge which task was harder and which was easier *before* making an epistemic inference. Most children answered these questions correctly—and yet, younger children still did not judge that the agent who succeeded on the more difficult diagnostic task was more likely to be knowledgeable.

While our results show that 6-year-olds can infer which agent is knowledgeable based on their relative chance of success, this experiment contrasted an event where success was assured against one where success was unlikely. Thus, children at this age might be able to qualitatively distinguish between assured and non-assured success, but not necessarily have the capacity to consider graded probabilities of success. We test this in Experiment 3, replicating Experiment 2 in a fully probabilistic setting.

5 | Experiment 3

5.1 | Method

5.1.1 | Participants

Sixty 5- to 6-year-olds (mean age: 5.93 years, range: 4.99–6.98 years) were recruited and tested online via the same video-chat research platform in the previous experiments. Because both 4- and 5-year-olds' performance did not exceed chance performance in the prior experiments on our primary test questions, we did not recruit 4-year-olds, preregistering a sample of only 5- and 6-year-olds. Seven additional participants were recruited but not included in the study (see Results). As before, we did not preregister collection of participant demographics. All data were collected between December 2022 and June 2023, with the exception of one data point that was collected in June 2024 to replace a participant who we accidentally failed to designate as "excluded" when stopping data collection, although they had failed inclusion.

5.1.2 | Stimuli

Stimuli were identical to that of Experiment 2, except we removed one marble from the "undiagnostic" side, so only three of the four boxes contained marbles; we now refer to this side as the "less diagnostic" side.

5.1.3 | Procedure

Figure 1 shows the experimental procedure. The procedure was identical to that of Experiment 2, but this time comparing a "less diagnostic" side with a 75% chance of random success to a "more diagnostic" side with a 25% chance of random success (see Figure 1). Success on the less diagnostic side was no longer assured, but the odds of random success were still substantially higher compared to the more diagnostic side (75% vs. 25% chance of random success).

5.2 | Results

For the 80.6% of participants whose sessions were video- or audio-taped ($n = 54/67$), two coders who were not involved in data collection determined exclusions according to preregistered criteria, as in the prior experiments. Seven participants were recruited but not included in the final sample because the participant did not provide codable answers to one or more test questions ($n = 2$), due to participant distraction/inattention ($n = 1$), experimenter error ($n = 1$), family interference ($n = 1$), connectivity issues ($n = 1$), or because the participant failed inclusion ($n = 1$).

Out of the final 60 participants included in the study, 65% of participants judged that the agent who succeeded on the more diagnostic task (finding the only marble on her side) had most likely peeked under the boxes. This proportion is reliably higher than chance ($n = 39$ of 60; 95% CI: 53.3–76.7). A logistic regression predicting performance as a function of centered age (continuous) did not reveal a significant effect of age ($\beta = 0.55$, $p = 0.24$). However, performance within each age group qualitatively differed. Only 60% of 5-year-olds ($n = 18$ of 30; 95% CI: 43.3–76.7) inferred that the agent who completed the more diagnostic task had prior knowledge, whereas 70% of 6-year-olds did so ($n = 21$ of 30; 95% CI: 53.3–86.7; see Figure 4).

Participants overall performed above chance when identifying where it would be more difficult to find a marble. 66.7% of participants ($n = 40$ of 60) correctly identified that it would be harder to find a marble on the more diagnostic side, a proportion reliably higher than chance (95% CI: 55–78.3). And 68.3% ($n = 41$ of 60) also correctly identified it would be easier to find a marble on the less diagnostic side, a proportion reliably higher than chance (95% CI: 56.7–80). A logistic regression predicting performance as a function of age did not reveal any significant effect of age for either test question (what's harder: $\beta = 0.02$, $p = 0.96$; what's easier: $\beta = -0.10$, $p = 0.84$). When looking at children's performance within each age group, 66.7% of 5-year-olds and 66.7% of 6-year-olds ($n = 20$ of 30, respectively; 95% CI: 50–83.3) correctly judged that it would be harder to find a marble on the diagnostic side (see Figure 4). Additionally, 66.7% of 5-year-olds ($n = 20$ of 30; 95% CI: 50–83.3) and 70% of 6-year-olds ($n = 21$ of 30; 95% CI: 53.3–86.7) judged that it would be easier to find a marble on the undiagnostic side (see Figure 4). See Supporting Information for explanations. Finally, a post hoc power analysis with a 70% pass rate (based upon 6-year-olds' performance on our test question, inferring who had peeked) yields power = 0.733.

5.3 | Discussion

These results replicate the developmental pattern observed in Experiments 1–2, in a fully probabilistic context: 6-year-olds judged that the agent who succeeded on the more diagnostic side—finding the *only* marble on her side—was more likely to have prior knowledge than the agent who succeeded in finding a marble on the less diagnostic side, where three of the four boxes contained marbles. 5-year-olds did not exceed chance performance, although their responses were in the right qualitative direction. These results suggest that 6-year-olds can indeed consider the probability of random success on each side, inferring that an agent who succeeds where random success is improbable must have had prior knowledge, whereas an agent who succeeds where random success is likely might have succeeded given ignorance.

Interestingly, although even infants discriminate proportions that differ by 50% (Denison and Xu 2014; Xu and Garcia 2008), participants in Experiment 3 seemed to experience qualitatively slightly increased difficulty identifying where it was

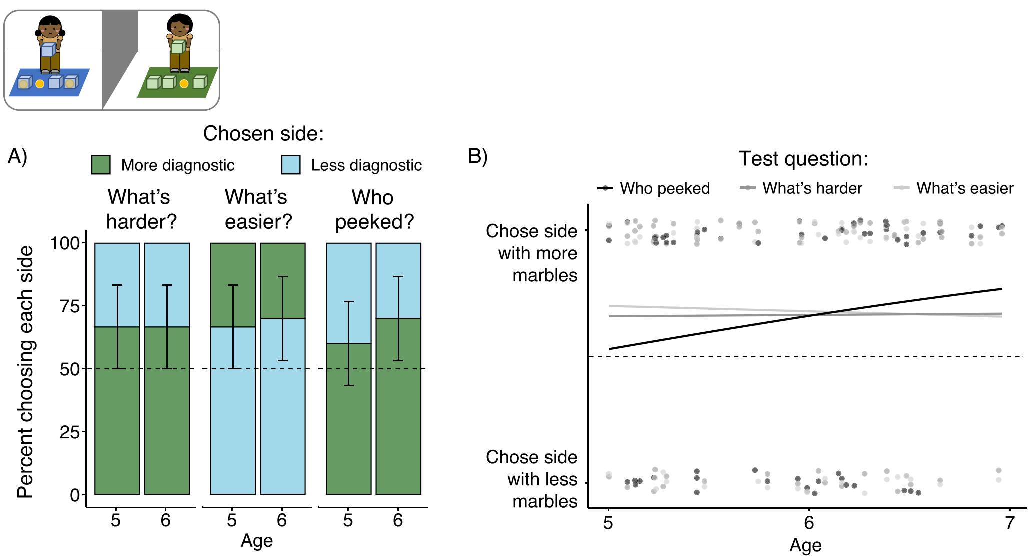

FIGURE 4 | At top left, see a basic procedure schematic. (A) Participant choices visualized by age group. The dotted line indicates predicted chance performance. Vertical bars show 95% bootstrapped confidence intervals. (B) Participant choices plotted continuously by age, along with a logistic regression fitted to each data set. Points are jittered along the *Y*-axis (but not the *X*-axis).

harder/easier to find a marble, with fewer participants identifying that random success was less probable (harder) on the side with only one marble and very probable (easier) on the side with three marbles. It is unclear whether these questions primarily tap participants' probability reasoning capacities per se versus their ability to represent and remember probabilities or to make a qualitative judgment about them (Doan et al. 2023).

6 | General Discussion

The capacity to teach or learn, help or hinder, and even punish or forgive relies at least in part on a capacity to infer what others know. However, inferring others' knowledge is complex; we must consider not only whether others failed or succeeded but also the probability they would have done so under different knowledge states. After all, even an ignorant agent is likely to succeed when the odds of random success are high. In mature Theory of Mind reasoning, inferences about other people's knowledge are not only sensitive to observed accuracy but also to the degree that this accuracy cannot be explained by chance alone. Here we found that this sensitivity does not emerge until age 6. Specifically, 6-year-olds, but not 4- and 5-year-olds, understood that asking someone to find a marble on a diagnostic side (where only one of four opaque boxes had a marble) would better reveal their knowledge, compared to an undiagnostic side (where every box had a marble). Conversely, when two agents found a marble in a set of boxes, only 6-year-olds consistently inferred that the agent who completed the diagnostic task (25% chance of success) was more likely to be knowledgeable, relative to the agent who completed the less diagnostic task (100% and 75% chance of success in Experiments 2 and 3, respectively). Interestingly, children of all ages were able to identify which task was harder, and asking this before or after the main task did not affect their epistemic reasoning. At first glance, 4- and 5-year-olds' difficulties may be surprising. After all, children can reason probabilistically from infancy (Denison et al. 2013; Xu and Garcia 2008; Gweon et al. 2010), and infer others' intentions and desires by considering the probability of their action outcomes from at least age 2 or 3 (Diesendruck et al. 2015; Kushnir et al. 2010; Ma and Xu 2011). But our results are consistent with a body of related work which suggests that an ability to integrate probability and belief to predict others' emotions is still developing throughout early childhood (Doan et al. 2018; MacLaren and Olson 1993; Ruffman and Keenan 1996; but see Scott 2017). For instance, 6-year-olds can accurately predict whether an agent will be surprised when asked to reason about the objective probability of an outcome if prompted to consider the outcome probability, but not when prompted to consider an agent's belief over this outcome (Doan et al. 2018).

The apparent divide in children's use of probability in mental-state reasoning is also 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 and Csibra 2003; Meltzoff 1995; Woodward 1998), it is not until age 4 or 5 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, 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 findings that a full understanding of epistemic states may continue to develop throughout the late preschool years (e.g., Aboody et al. 2019; Wu and Schulz 2018). For instance, it is not until age 6 that children even begin to appreciate that ignorant agents will search randomly (Ruffman 1996; Chen et al. 2015; Friedman and Petrashek 2009). If 4- and 5-year-olds in our task did not expect that an ignorant agent would search randomly for a marble, this could explain why younger children did not prefer to ask about—and did not infer knowledge from success on—the diagnostic task. 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 epistemic states like knowledge and ignorance (see also Phillips et al. 2021).

One limitation of our study is that our experiments were not powered to detect small effect sizes. This raises the possibility that, with a larger sample size, 5-year-olds' performance would have been considered reliably above chance. Indeed, 5-year-olds' test question responses consistently trended in the right direction, with 56.7%, 63.3%, and 60% sharing 6-year-olds' judgments across Experiments 1, 2, and 3, respectively. This suggests that about 60% of 5-year-olds might be answering in a mature way. However, even if a larger sample size allowed us to establish that this preference is reliably higher than chance, this would not affect our general conclusions. After all, this weak effect size would suggest that many 5-year-olds still lack this intuition and still points to our conclusion that children begin to integrate probabilistic reasoning into their knowledge inferences at around age 5 or 6.

Another limitation of our work is that we did not preregister the collection of demographic details in our experiments. This is because there is little evidence that basic Theory of Mind capacities vary meaningfully across race/ethnicity or gender (see Doebel and Frank 2024); in addition, online recruitment methods likely reflect more representative demographics than in-person research, which samples primarily from areas surrounding an elite institution (see Sheskin et al. 2020). Nevertheless, our sample was North American, and we do not know to what extent this developmental timeline would vary across cultures. In the same way that other aspects of Theory of Mind development vary cross-culturally (Wellman et al. 2001), it is possible that children's exact age at which they begin to integrate probabilistic reasoning might also appear earlier or later in development in different cultures.

While our work established a Theory of Mind limitation in young children, it did not reveal why younger children fail our task, or what causes children to shift to a probabilistic understanding of accuracy and knowledge. While this is a direction for future work to explore, there are at least two hypotheses that could explain this data. To start, children's success in our task coincides with their success in second-order false belief tasks (Perner and Wimmer 1985). This ability to reason about nested mental states might be critical for Experiment 1, where children might need to ask themselves "What kind of information would I need to figure out what the protagonist knows?" While this kind of reasoning is not required in Experiments 2 and 3 (where children directly observe agents' actions and their outcomes), children in these experiments might face a different challenge: representing two different minds with different epistemic states. Thus, children's overall pattern of failures might be explained due to general limitations representing nested mental states and mental states of different agents at once. Note, however, that many experiments with 4-year-olds ask children to represent two agents' mental states at once, and they can generally do so easily, challenging this idea (e.g., Aboody et al. 2022; Chuey, Jara-Ettinger, et al. 2023; Kushnir et al. 2013; Lutz and Keil 2002; Morris et al. (n.d.); Jara-Ettinger et al. 2017).

Another possibility comes from constructivist theories of Theory of Mind (e.g., Goodman et al. 2006; Ullman and Tenenbaum 2020; Xu 2019), where children "upgrade" their mental representations in response to evidence of their explanatory power. Under this view, children might begin with a simple model that treats accuracy and knowledge as equivalent, which is an efficient way to track knowledge in a way that gets things broadly right. As children enter schooling and start to face a variety of informants, including unreliable ones, children might notice conflicts in their reasoning. For instance, children might acquire enough evidence to be confident that some of their peers are not particularly knowledgeable, while also seeing them be accurate at a variety of simple tasks. In attempting to explain how this is the case, children might become more attuned to the role of difficulty and transition to a richer model where accuracy and knowledge are mediated by the probability of being right by chance. Some evidence for this possibility comes from the fact that children at this age can already reason about difficulty in the context of mental states (Heyman et al. 2003), and performed near ceiling in this question in all our tasks. This suggests that children might already have the conceptual building blocks necessary to succeed in our tasks, but have not yet integrated them into their model.

While our results show that 5-year-olds are not integrating probabilistic reasoning into epistemic inferences, there is a question of how exactly 6-year-olds succeeded. Is it possible that they relied on a simpler cue-based strategy that does not count as full integration? For instance, children might learn to associate lower baseline probabilities of success as a cue to higher degrees of knowledge, without having a causal understanding that expresses the relationship between mental states and behavior in probabilistic terms. We believe this is unlikely because children at this age are already building a coherent causal model of how mental states relate to behavior (Aboody et al. 2022; Jara-Ettinger et al., 2016), and these models are already probabilistic (Ullman and Tenenbaum 2020). However, this is an open question. If 6-year-olds are indeed using an upgraded cue-based judgment, this would imply there is a second, later stage at which children transition to integrate probabilistic reasoning into their mental model of other minds.

Despite these open questions, our work makes two direct contributions to debates in the literature. First, a popular theory of social learning has argued that children must have epistemic vigilance (Harris et al. 2018; Sperber et al. 2010). That is, social learning is a powerful way to transmit complex knowledge, but it comes at the risk of receiving information that can be accidentally or intentionally misleading. Our work suggests that young children are still developing their ability to determine who is knowledgeable and whom to trust, despite often making correct inferences and decisions. This possibility is consistent with common observations about early learning. While children learn a lot of accurate facts about the world from others, they also trustfully believe fictions that are inconsistent with everything else they know, such as the existence of Santa Claus, that babies are delivered by storks, or that your nose will noticeably grow when you lie. The evidence we present here, revealing some limits on children's abilities to determine who is knowledgeable, lends support to the idea that this capacity might best be thought of as one that is in a period of intense development, rather than a complex one that is already playing a foundational role in supporting early social reasoning.

Our work also contributes to the question of what exactly changes in children's understanding of other minds. Research in the last 20 years has been characterized by the discovery of sophisticated Theory of Mind reasoning from early in infancy (e.g., Onishi and Baillargeon 2005; Liu et al. 2017), leaving open the question of why young children's social skills continue to appear limited relative to adults. One major source of this difference lies in domain-general maturation: namely, children's developing executive functions, which children might need to be able to suppress their own beliefs when answering questions about other minds. However, beyond the domain-general maturation children undergo during the preschool years, our work adds to the literature suggesting that children's domain-specific understanding of how minds work is also changing (Wellman et al. 2001; Wellman and Liu 2004).

Our work also opens the possibility that there may be further areas of development within children's epistemic reasoning. Specifically, here we focused on one of the simplest intuitions that combines probability of success and knowledge. But these intuitions are even more complex in adults. 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 did not know they would win). This suggests that adults do not treat low-probability success as a context-insensitive cue to knowledge. We instead recognize that unlikely events do indeed happen, and we might be more willing to accept that when it is difficult to conceive how someone could come to acquire the relevant knowledge. Moreover, in real-world situations, estimating how unlikely it might be for someone to succeed in a task is a challenging problem and therefore, even after age 6, children's inferences might be limited by their developing understanding of how to estimate probabilistic relations in complex events. We hope to explore these dimensions in future work.

7 | Conclusion

To infer the cause of a failed action, figure out what to teach, or decide who knew better, we must understand others' epistemic states. In the current work, we find that by age 6, 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 Madison Flowers, Mackenzie Briscoe, Colin Jacobs, Lauren Barragan, Jenna Landy, Nathalia Reis, Hudson Patterson, Sofia Rubio, Eden Senay, Caroline Telesz, Gamze Kazakoglu, and Janice Dean for assistance with data collection. We thank Ilayda Orhan, Rodney Tompkins, Nathalia Reis, Janice Dean, Joshua Bolchover, Mackenzie Briscoe, and Areej Rizvi for assistance with coding. We thank Katie Vasquez for her help with participant recruitment. We thank an anonymous reviewer for pointing out the relation to second-order false belief reasoning. This work was supported by NSF award BCS-2045778 to J.J.E. and a Franke Interdisciplinary Graduate Award from Yale's Franke Program in Science and the Humanities to R.A. The procedures, predictions, sample size, exclusion criteria, and analysis plan for all experiments were pre-registered and are publicly available on the project OSF page in addition to all data and code necessary to reproduce analyses and all materials necessary to attempt to replicate the findings. Link: https://osf.io/fm8e2.

Data Availability Statement

The procedures, predictions, sample size, exclusion criteria, and analysis plan for all experiments were pre-registered, and are publicly available in the project OSF page in addition to all data and code necessary to reproduce analyses, and all materials necessary to attempt to replicate the findings. Link: https://osf.io/fm8e2.

References

Aboody, R., H. Huey, and J. Jara-Ettinger. 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., M. Flowers, C. Zhou, and J. Jara-Ettinger. 2019. "Ignorance = Doing What Is Reasonable: Children Expect Ignorant Agents to Act Based on Prior Knowledge." In *Proceedings of the 41st Annual Conference of the Cognitive Science Society*, 1297–1303. Cognitive Science Society.

Bridgers, S., D. Buchsbaum, E. Seiver, T. L. Griffiths, and A. Gopnik. 2016. "Children's Causal Inferences From Conflicting Testimony and Observations." *Developmental Psychology* 52, no. 1: 9–18.

Bridgers, S., J. Jara-Ettinger, and H. Gweon. 2020. "Young Children Consider the Expected Utility of Others' Learning to Decide What to Teach." *Nature human behaviour* 4, no. 2: 144–152.

Brosseau-Liard, P. E., and S. A. Birch. 2010. "'I Bet You Know More and Are Nicer Too!': What Children Infer From Others' Accuracy." *Developmental Science* 13, no. 5: 772–778.

Chen, Y., Y. Su, and Y. Wang. 2015. "Young Children Use the "Ignorance=Getting It Wrong" Rule When Predicting Behavior." *Cognitive Development* 35: 79–91.

Chuey, A., J. Jara-Ettinger, and H. Gweon. 2023. "Violation of Epistemic Expectations: Children Monitor What Others Know and Recognize Unexpected Sources of Knowledge." In *Proceedings of the 45th Annual Conference of the Cognitive Science Society*. Cognitive Science Society.

Chuey, A., R. Sparks, and H. Gweon. 2023. "Young Children Can Identify Knowledgeable Speakers from their Causal Influence Over Listeners." *In Proceedings of the Annual Meeting of the Cognitive Science Society* 45, no. 45.

Cohen, J. 1994. "The Earth Is Round (p<. 05)." *American Psychologist* 49, no. 12: 997, The earth is round ($p < .05$)–1003.

Corriveau, K., and P. L. Harris. 2009. "Preschoolers Continue to Trust a More Accurate Informant 1 Week After Exposure to Accuracy Information." *Developmental Science* 12, no. 1: 188–193.

Crivello, C., and D. Poulin-Dubois. 2018. "Infants' False Belief Understanding: A Non-Replication of the Helping Task." *Cognitive Development* 46: 51–57.

Cumming, G. 2014. "The New Statistics: Why and How." *Psychological Science* 25, no. 1: 7–29.

Denison, S., and F. Xu. 2014. "The Origins of Probabilistic Inference in Human Infants." *Cognition* 130, no. 3: 335–347.

Denison, S., C. Reed, and F. Xu. 2013. "The Emergence of Probabilistic Reasoning in Very Young Infants: Evidence From 4.5- and 6-Month-Olds." *Developmental Psychology* 49, no. 2: 243–249.

Diesendruck, G., S. Salzer, T. Kushnir, and F. Xu. 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, no. 2: 370–380.

Doan, T., O. Friedman, and S. Denison. 2018. "Beyond Belief: The Probability-Based Notion of Surprise in Children." *Emotion* 18, no. 8: 1163–1173.

Doan, T., O. Friedman, and S. Denison. 2023. "Calculated Feelings: How Children Use Probability to Infer Emotions." *Open Mind* 7: 879–893.

Doebel, S., and M. C. Frank. 2024. "Broadening Convenience Samples to Advance Theoretical Progress and Avoid Bias in Developmental Science." *Journal of Cognition and Development* 25, no. 2: 261–272.

Einav, S., and E. J. Robinson. 2011. "When Being Right Is Not Enough: Four-Year-Olds Distinguish Knowledgeable Informants From Merely Accurate Informants." *Psychological Science* 22: 1250–1253.

Fabricius, W. V., C. R. Gonzales, A. Pesch, et al. 2021. "Perceptual Access Reasoning (PAR) in Developing a Representational Theory of Mind." *Monographs of the Society for Research in Child Development* 86, no. 3: 7–154.

Friedman, O., and A. R. Petrashek. 2009. "Children Do Not Follow the Rule "Ignorance Means Getting It Wrong"." *Journal of Experimental Child Psychology* 102, no. 1: 114–121.

Gergely, G., and G. Csibra. 2003. "Teleological Reasoning in Infancy: The Naïve Theory of Rational Action." *Trends in Cognitive Sciences* 7: 287–292.

German, T. P., and A. M. Leslie. 2001. "Children's Inferences From 'Knowing' to 'Pretending' and 'Believing'." *British Journal of Developmental Psychology* 19, no. 1: 59–83.

Goodman, N. D., C. L. Baker, E. B. Bonawitz, et al. 2006. "Intuitive Theories of Mind: A Rational Approach to False Belief." In *Proceedings of the Twenty-Eighth Annual Conference of the Cognitive Science Society*, vol. 6. Cognitive Science Society.

Gweon, H., J. B. Tenenbaum, and L. E. Schulz. 2010. "Infants Consider Both the Sample and the Sampling Process in Inductive Generalization." *Proceedings of the National Academy of Sciences* 107, no. 20: 9066–9071.

Harris, P. L. 2012. *Trusting What You're Told: How Children Learn From Others*. Harvard University Press.

Harris, P. L., M. A. Koenig, K. H. Corriveau, and V. K. Jaswal. 2018. "Cognitive Foundations of Learning From Testimony." *Annual Review of Psychology* 69, no. 1: 251–273.

Heyman, G. D., C. L. Gee, and J. W. Giles. 2003. "Preschool Children's Reasoning About Ability." *Child Development* 74, no. 2: 516–534.

Jara-Ettinger, J., S. Floyd, J. B. Tenenbaum, and L. E. Schulz. 2017. "Children Understand That Agents Maximize Expected Utilities." *Journal of Experimental Psychology: General* 146: 1574–1585.

Kampis, D., P. Karman, G. Csibra, V. Southgate, and M. Hernik. 2021. "A Two-Lab Direct Replication Attempt of Southgate, Senju and Csibra (2007)." *Royal Society Open Science* 8, no. 8: 210190.

Koenig, M. A., and P. L. Harris. 2005. "Preschoolers Mistrust Ignorant and Inaccurate Speakers." *Child Development* 76, no. 6: 1261–1277.

Koenig, M. A., F. Clément, and P. L. Harris. 2004. "Trust in Testimony: Children's Use of True and False Statements." *Psychological Science* 15: 694–698.

Kushnir, T., C. Vredenburgh, and L. A. Schneider. 2013. ""Who Can Help Me Fix This Toy?" the Distinction Between Causal Knowledge and Word Knowledge Guides Preschoolers' Selective Requests for Information." *Developmental Psychology* 49, no. 3: 446–453.

Kushnir, T., F. Xu, and H. M. Wellman. 2010. "Young Children Use Statistical Sampling to Infer the Preferences of Other People." *Psychological Science* 21, no. 8: 1134–1140.

Liu, S., T. D. Ullman, J. B. Tenenbaum, and E. S. Spelke. 2017. "Ten-Month-Old Infants Infer the Value of Goals from the Costs of Actions." *Science* 358, no. 6366: 1038–1041.

Lutz, D. J., and F. C. Keil. 2002. "Early Understanding of the Division of Cognitive Labor." *Child Development* 73, no. 4: 1073–1084.

Ma, L., and F. Xu. 2011. "Young Children's Use of Statistical Sampling Evidence to Infer the Subjectivity of Preferences." *Cognition* 120, no. 3: 403–411.

MacLaren, R., and D. Olson. 1993. "Trick or Treat: Children's Understanding of Surprise." *Cognitive Development* 8, no. 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, no. 5: 838–850.

Morris, B., D. Yurovsky, and A. Shaw. n.d (invited revision). "Um..." Thinking out Loud: Children Infer Social Meaning From Speech Disfluencies."

Nurmsoo, E., and E. J. Robinson. 2009. "Identifying Unreliable Informants: Do Children Excuse Past Inaccuracy?" *Developmental Science* 12: 41–47.

O'Neill, D. K. 1996. "Two-Year-Old Children's Sensitivity to a Parent's Knowledge State When Making Requests." *Child Development* 67, no. 2: 659–677.

Onishi, K. H., and R. Baillargeon. 2005. "Do 15-Month-Old Infants Understand False Beliefs?" *Science* 308, no. 5719: 255–258.

Pasquini, E. S., K. H. Corriveau, M. Koenig, and P. L. Harris. 2007. "Preschoolers Monitor the Relative Accuracy of Informants." *Developmental Psychology* 43, no. 5: 1216–1226.

Perner, J., and H. Wimmer. 1985. "John Thinks that Mary Thinks that..." Attribution of Second-Order Beliefs by 5-to 10-Year-Old Children." *Journal of Experimental Child Psychology* 39, no. 3: 437–471.

Phillips, J., W. Buckwalter, F. Cushman, et al. 2021. "Knowledge Before Belief." *Behavioral and Brain Sciences* 44: e140.

Ruffman, T. 1996. "Do Children Understand the Mind by Means of Simulation or a Theory? Evidence From Their Understanding of Inference." *Mind & Language* 11, no. 4: 388–414.

Ruffman, T., and T. R. Keenan. 1996. "The Belief-Based Emotion of Surprise: The Case for a Lag in Understanding Relative to False Belief." *Developmental Psychology* 32, no. 1: 40.

Saxe, R. 2005. "Against Simulation: The Argument From Error." *Trends in Cognitive Sciences* 9, no. 4: 174–179.

Schuwerk, T., B. Priewasser, B. Sodian, and J. Perner. 2018. "The Robustness and Generalizability of Findings on Spontaneous False Belief Sensitivity: A Replication Attempt." *Royal Society Open Science* 5, no. 5: 172273.

Scott, R. M. 2017. "Surprise! 20-Month-Old Infants Understand the Emotional Consequences of False Beliefs." *Cognition* 159: 33–47.

Sheskin, M., K. Scott, C. M. Mills, et al. 2020. "Online Developmental Science to Foster Innovation, Access, and Impact." *Trends in Cognitive Sciences* 24, no. 9: 675–678.

Sperber, D., F. Clément, C. Heintz, et al. 2010. "Epistemic vigilance." *Mind & Language* 25, no. 4: 359–393.

Ting, F., Z. He, and R. Baillargeon. 2021. "Five-Month-Old Infants Attribute Inferences Based on General Knowledge to Agents." *Journal of Experimental Child Psychology* 208: 105126.

Ullman, T. D., and J. B. Tenenbaum. 2020. "Bayesian Models of Conceptual Development: Learning as Building Models of the World." *Annual Review of Developmental Psychology* 2, no. 1: 533–558.

Warneken, F., and M. Tomasello. 2006. "Altruistic Helping in Human Infants and Young Chimpanzees." *Science* 311, no. 5765: 1301–1303.

Wellman, H. M. 2011. "Reinvigorating Explanations for the Study of Early Cognitive Development." *Child Development Perspectives* 5, no. 1: 33–38.

Wellman, H. M. 2014. *Making Minds: How Theory of Mind Develops*. Oxford University Press.

Wellman, H. M., and D. Liu. 2004. "Scaling of theory-of-mind tasks." *Child Development* 75, no. 2: 523–541.

Wellman, H. M., D. Cross, and J. Watson. 2001. "Meta-Analysis of Theory-Of-Mind Development: The Truth About False Belief." *Child Development* 72, no. 3: 655–684.

Wellman, H. M., S. Carey, L. Gleitman, E. L. Newport, and E. S. Spelke. 1990. *The Child's Theory of Mind*. MIT Press.

Woodward, A. L. 1998. "Infants Selectively Encode the Goal Object of an Actor's Reach." *Cognition* 69, no. 1: 1–34.

Wu, Y., and L. E. Schulz. 2018. "Inferring Beliefs and Desires From Emotional Reactions to Anticipated and Observed Events." *Child Development* 89, no. 2: 649–662.

Xu, F. 2019. "Towards a Rational Constructivist Theory of Cognitive Development." *Psychological Review* 126, no. 6: 841–864.

Xu, F., and V. Garcia. 2008. "Intuitive Statistics by 8-Month-Old Infants." *Proceedings of the National Academy of Sciences* 105: 5012–5015.

Young, L., F. Cushman, M. Hauser, and R. Saxe. 2007. "The Neural Basis of the Interaction Between Theory of Mind and Moral Judgment." *Proceedings of the National Academy of Sciences* 104, no. 20: 8235–8240.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.