Changing minds: Children's inferences about third party belief revision

Received: 11 December 2015 | Accepted: 5 December 2016

DOI: 10.1111/desc.12553

PAPER

WILEY Developmental Science

Changing minds: Children's inferences about third party belief revision

Rachel W. Magid[1,*] | Phyllis Yan[1,2,*] | Max H. Siegel[1] | Joshua B. Tenenbaum[1] | Laura E. Schulz[1]

[1]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

[2]Department of Statistics, University of Michigan, Ann Arbor, Michigan, USA

Correspondence
Rachel Magid, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 77 Massachusetts Ave, 46-4011, Cambridge, MA 02139, USA
Email: rwmagid@mit.edu

Funding information
Division of Computing and Communication Foundations, Grant/Award Number: 1231216; Division of Research on Learning in Formal and Informal Settings, Grant/Award Number: 0744213; Center for Brains, Minds, and Machines (CBMM); NSF STC, Grant/Award Number: CCF-1231216; National Science Foundation, Grant/Award Number: 0744213

Abstract

By the age of 5, children explicitly represent that agents can have both true and false beliefs based on epistemic access to information (e.g., Wellman, Cross, & Watson, 2001). Children also begin to understand that agents can view identical evidence and draw different inferences from it (e.g., Carpendale & Chandler, 1996). However, much less is known about when, and under what conditions, children expect other agents to change their minds. Here, inspired by formal ideal observer models of learning, we investigate children's expectations of the dynamics that underlie third parties' belief revision. We introduce an agent who has prior beliefs about the location of a population of toys and then observes evidence that, from an ideal observer perspective, either does, or does not justify revising those beliefs. We show that children's inferences on behalf of third parties are consistent with the ideal observer perspective, but not with a number of alternative possibilities, including that children expect other agents to be influenced only by their prior beliefs, only by the sampling process, or only by the observed data. Rather, children integrate all three factors in determining how and when agents will update their beliefs from evidence.

RESEARCH HIGHLIGHTS

- Understanding the conditions under which other agents will change their minds is a key component of social cognition.
- Considerable evidence suggests that children themselves learn rationally from data: integrating evidence with their prior beliefs.
- Do 4- to 6-year-olds expect other agents to learn rationally? Can they use others' prior beliefs and data to predict when third parties will retain their beliefs and when they will change their minds?
- Here we use a computational model of rational learning to motivate predictions for an ideal observer account, as well as five alternative accounts. We found that children expect third parties to be rational learners with respect to their own prior beliefs.
- The data were not consistent with alternative accounts. In particular, children did not expect others simply to retain their own prior beliefs, learn from the data without integrating it with their prior beliefs, or share the children's beliefs. Rather, children expected agents to learn normatively from evidence.

1 | INTRODUCTION

Expectations of rational agency support our ability to predict other people's actions and infer their mental states (Dennett, 1987; Fodor, 1987). Adults assume that agents will take efficient routes towards their goals (D'Andrade, 1987; Heider, 1958), and studies with infants suggest that these expectations emerge very early in development

[*] The first two authors contributed equally to this work.

(Skerry, Carey, & Spelke, 2013). By the end of the first year, infants can use situational constraints, along with knowledge about an agent's goal, to predict an agent's actions. Similarly, they use knowledge of an agent's actions and situational constraints to infer the agent's goal, as well as knowledge of an agent's actions and goal to infer unobserved situational constraints (Csibra, Bíró, Koós, & Gergely, 2003; Gergely & Csibra, 2003; Gergely, Nádasdy, Csibra, & Bíró, 1995). Such work has inspired computational models of theory of mind that formalize the principle of rational action and successfully predict human judgments (Baker, Saxe, & Tenenbaum, 2009; Baker, Saxe, & Tenenbaum, 2011; Jara-Ettinger, Baker, & Tenenbaum, 2012). Here however, we ask whether learners' expectations extend to the more colloquial meaning of the word 'rational': the expectation that other people's judgments and beliefs have a basis in the evidence they observe.

Note that this is distinct from the question of whether children *themselves* draw rational inferences from data. Decades of research suggest that very young children can integrate prior beliefs with small samples of evidence to infer the extensions of word meanings, identify object categories, learn causal relationships, and reason about others' goal-directed actions (see Gopnik & Wellman, 2012; Schulz, 2012; and Tenenbaum, Kemp, Griffiths, & Goodman, 2011, for reviews). However, despite extensive work on children's theory of mind (see Wellman, 2014, for discussion and review), less is known about how children expect others to learn from evidence. Although classic theory of mind tasks look at whether children expect others to update their beliefs given diverse forms of epistemic access to data – including direct perceptual access (e.g., Wimmer & Perner, 1983), indirect clues (e.g., Sodian, Taylor, Harris, & Perner, 1991) and testimony (e.g., Zaitchik, 1991) – these involve a relatively simple instantiation of the expectation that others will learn based on their observations of the world: children need only understand whether the agent does, or does not, have epistemic access to belief-relevant information. Such studies do not ask whether children understand that agents might evaluate evidence differently or draw different inferences from identical evidence.

The studies that do look at children's understanding of how third parties might evaluate evidence suggest that an 'interpretative theory of mind' is a relatively late development (Astington, Pelletier, & Homer, 2002; Carey & Smith, 1993; Chandler & Carpendale, 1998; LaLonde & Chandler, 2002; Myers & Liben, 2012; Pillow & Mash, 1999; Ross, Recchia, & Carpendale, 2005; Ruffman, Perner, Olson, & Doherty, 1993). Not until 6 and 7 years do children understand, for example, that an ambiguous line drawing can be viewed as two different kinds of animals (Carpendale & Chandler, 1996) or that iconic symbols are subject to different interpretations (Myers & Liben, 2012). Young children's failure to understand that agents can reach different conclusions from the same evidence suggests that children might have difficulty understanding how other agents' prior knowledge affects the interpretation of data.

Arguably, however, understanding that evidence is ambiguous and thus open to interpretation may be more challenging than understanding the conditions under which others might be expected to learn from evidence. Relatively little work has looked at what children understand about others' inferences from data, and the findings here are mixed. For instance, both 4- and 6-year-olds recognize that an unseen marble must be blue if it is drawn from a bag containing only blue marbles; however, only 6-year-olds recognize that a third party (who knows the contents of the bag) will make the same inference and thus know the color of the marble (Sodian & Wimmer, 1987). However, 4-year-olds do understand that if covariation evidence suggests that one of two causes is correlated with an outcome and the experimenter tricks a puppet by reversing the evidence, the puppet will conclude that the wrong variable is the cause (Ruffman et al., 1993).

Such studies suggest that by 4, children are at least beginning to understand that third parties learn from evidence in ways that go beyond mere perceptual access to data. However, they leave open the question of whether children can use patterns of evidence to understand when others will change their minds, and the degree to which children integrate others' prior beliefs in predicting their learning. Do children expect others to update their beliefs from data in cases where learning requires representing not merely an agent's access to evidence but the agent's ability to draw appropriate inferences from the evidence?

To ask whether children expect others to rationally update their beliefs from data we borrow from two influential tasks in the literature. The first is the classic false belief task (Wimmer & Perner, 1983). The other is derived from work looking at infants' and children's understanding of the relationship between samples and populations (e.g., Denison & Xu, 2014; Gweon, Tenenbaum, & Schulz, 2010; Kushnir, Xu, & Wellman, 2010; Xu & Denison, 2009; Xu & Garcia, 2008). Specifically, we show a child and another agent (a Frog puppet) two boxes: one containing more rubber ducks than ping-pong balls (the Duck box) and one containing more balls than ducks (the Ball box). The Frog leaves, and the child watches as the boxes are either moved and returned to the same location (so the Frog has a true belief about the location of each box) or switched (so the Frog has a false belief about the location of each box). At test, the Frog returns, and both the child and the Frog watch as the experimenter reaches into the Duck box and draws a sample of three or five ducks either apparently at random (without looking into the box) or selectively (looking in and fishing around). After both the child and the Frog see the sample of data, children are asked, 'Where does Froggy think the Duck box is now?' See Figure 1 for a schematic of the procedure.

Both the ability to reason about others' false beliefs (see Baillargeon, Scott, & He, 2010) and the ability to recognize when data are sampled randomly or selectively (e.g., Xu & Denison, 2009) emerge relatively early in development. However, children do not reliably provide accurate responses in explicit false belief tasks until later childhood (see Wellman, Cross, & Watson, 2001, for review) and as noted, the ability to understand that identical evidence can be open to different interpretations emerges even later (e.g., Astington et al., 2002; Carey & Smith, 1993, Chandler & Carpendale, 1998; LaLonde & Chandler, 2002; Myers & Liben, 2012; Pillow & Mash, 1999; Ross et al, 2005; Ruffman et al., 1993). Because we are interested not in children's own inferences from the data, but in their inferences on behalf of a third party whose beliefs may differ both from the child's own and those supported by the observed data, here we focus on 4.5- to 6-year-olds.

As shown in Table 1, if children expect the Frog to update his beliefs from evidence, then the cross between old and new locations

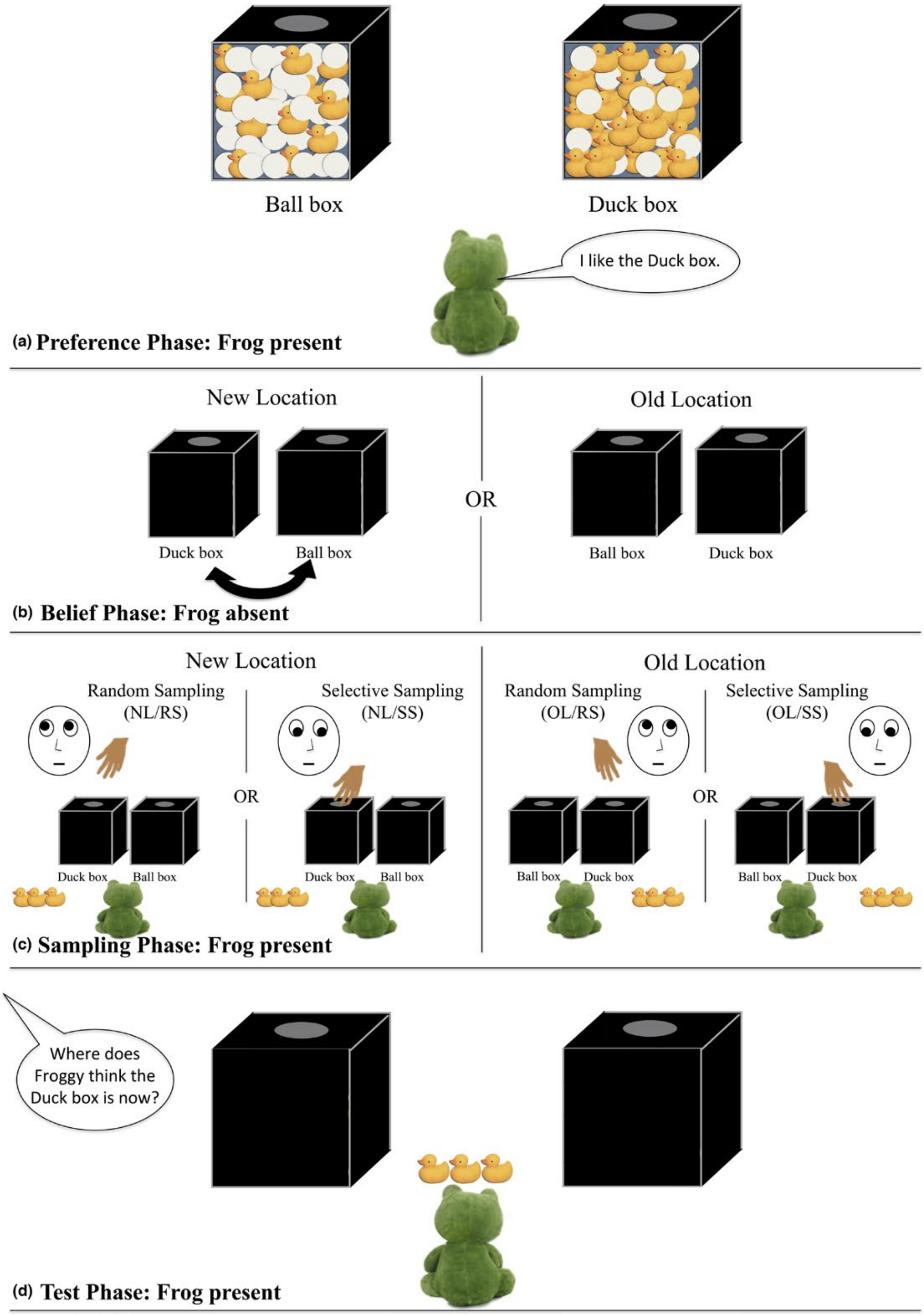


FIGURE 1 Schematic of the procedure. In the Preference phase (a) children are shown the two boxes with different proportions of ducks and balls and asked to identify the Duck box and Ball box based on each box's majority object. Then children are introduced to the Frog puppet and his preference for ducks and the Duck box and then learn, along with the Frog, that the boxes can either each move back and forth to stay in the same location or move from one side to the other to switch locations. In the Belief Phase (b) children either see the boxes switch locations (New Location condition) or stay in the same location (Old Location condition) while the Frog is absent. When the Frog returns, he will either have a false belief about the location of the Duck box (New Location condition) or a true belief about the location of the Duck box (Old Location condition). Children are asked two check questions to confirm that they have tracked the locations of the boxes and the Frog's belief at the end of the Belief Phase. In the Sampling Phase (c) the Frog returns and the experimenter samples either randomly (Random Sampling condition) or selectively (Selective Sampling condition) from the hidden Duck box. At the Test Phase children are asked where the Frog thinks the Duck box is

and random and selective sampling predicts a pattern of responses distinct from the pattern that would be generated if children adopted many other possible response strategies. We will walk through the predictions of our account intuitively; however, to clarify our proposal, we also include a computational model providing quantitative predictions for both our account and a number of alternatives, in each experimental condition (see Figure 4 and Appendix S1). The details of the model are not critical to our proposal as our goal here is not to evaluate the Rational Learning model per se. Given that there are only five conditions, and some of them (e.g., both selective sampling conditions) make overlapping predictions, correlations between the model and children's performance may be less convincing than the relative fit of the Rational Learning model in comparison to the alternative models. That is the analysis we include here. In addition, it is helpful to consider the qualitative intuitions behind these models insofar as they motivate our predictions and ground our intuitions in a precise statement of what constitutes 'rational inference' in this context.

1.1 | Predictions of the rational inference account

If children expect agents to rationally update their beliefs, they should respond jointly to the type of sampling process and the Frog's prior beliefs about the boxes' locations together with his knowledge that the boxes can move. A sample randomly drawn from a population is likely to be representative of the population. Thus we predict that in the Random Sampling conditions, children should expect the Frog to use the evidence to verify or update his beliefs about the location of the Duck box.

Specifically, randomly sampling three ducks in a row is improbable unless the evidence is sampled from the Duck box. Thus when evidence is randomly sampled from the Old Location (OL/RS), children should infer that the Frog will retain his belief and will continue to think that the Duck box is in the Old location. However, when evidence is randomly sampled from the New Location (NL/RS_3 ducks and NL/RS_5 ducks) children should believe that the Frog may now update his former false belief, inferring that the Duck box may have been moved to the New Location. Moreover, the strength of children's inferences should depend, in a graded way, on the strength of evidence they observe: they should be more confident that the Frog will change his mind when they see five ducks randomly drawn from the New Location (NL/RS_5 ducks) than when they see three ducks (NL/RS_3 ducks) randomly drawn.

By contrast, selectively sampled evidence is uninformative about the population from which it is drawn. The experimenter can selectively draw any sample at all (representative or non-representative) from the population. Indeed, if the experimenter is trying to *guarantee* that she gets three ducks in a row, she should sample selectively regardless of whether she is drawing from the population where ducks are relatively common (the Duck box) or the population where ducks are relatively rare (the Ball box). Since the selectively sampled evidence is consistent with sampling from either box, a rational learner who integrates his prior beliefs with the data should retain his prior beliefs. That is, because both the Old and New Location are consistent with the data and only the Old Location is consistent with the agent's prior beliefs, we predict that children will expect the Frog to say the Duck box is in the Old Location in both the Old (OL/SS) and New Location (NL/SS) Selective Sampling conditions.

1.2 | Alternative accounts

In contrast to the pattern of responses consistent with third party rational inference (Table 1, row a), there are a number of other ways children might respond to the question 'Where does Froggy think the Duck box is now?' Children might respond with the actual location of

TABLE 1 The predictions for the dominant response pattern if children expect other agents to engage in rational learning from data are listed in row a. The * indicates that the probability that children think the Frog will change his mind should depend on the strength of the evidence the Frog observes. Possible alternative patterns of responses to the test question in each of the four conditions: New Location/Random Sampling (NL/RS); New Location/Selective Sampling (NL/SS); Old Location/Random Sampling (OL/RS); Old Location/Selective Sampling (OL/SS). OLD indicates that the child would point to the original location of the Duck box and NEW that the child would point to the new location

Response pattern	New Location Random Sampling NL/RS	New Location Selective Sampling NL/SS	Old Location Random Sampling OL/RS	Old Location Selective Sampling OL/SS
a. Rational Learning	NEW*	OLD	OLD	OLD
b. Actual location (or child's own beliefs)	NEW	NEW	OLD	OLD
c. Frog's beliefs (without updating from data)	OLD	OLD	OLD	OLD
d. Sampled data (without prior beliefs)	NEW	CHANCE	OLD	CHANCE
e. Random-Stay; Selective-Shift	NEW	OLD	OLD	NEW
f. Chance	CHANCE	CHANCE	CHANCE	CHANCE

the Duck box; this is, after all, the location from which the Frog sees the ducks sampled, and also of course consistent with what the children themselves believe (row b, Table 1). Alternatively, children might respond with the Frog's true or false belief about the location of the Duck box, considering whether the Frog saw the boxes moved or not but without expecting Frog to update his beliefs given the sampled evidence (row c, Table 1). Another possibility is that children might expect the Frog to attend to the sampled evidence but not integrate it with his prior beliefs; they may conclude that if the Frog sees randomly sampled evidence he will strongly conclude that the Ducks are in that location, but if he sees selectively sampled evidence, he will recognize that the evidence is uninformative and choose at chance (row d, Table 1). Yet another possibility is that the children think the Frog will attend to the sampled evidence but not as a rational learner would; they might, for instance, think that Frog will conclude that random sampling indicates that the sample of ducks is pulled from the Duck box and that selective sampling of ducks means the sample is pulled from the Ball box (row e, Table 1). Finally, children might respond at chance, either because they genuinely believe that the Frog will guess or because different children choose different strategies and thus, as a group, generate responses indistinguishable from chance responding (row f, Table 1). Corresponding to the qualitative predictions shown in Table 1, Figure 4 shows quantitative predictions for the rational inference account and each of the alternative accounts for all of the conditions in our study. The model predictions can be compared with children's behavioral data.

In the experiment to follow, we test these different accounts and predict that children's responses will be best explained as inferring that the Frog will rationally integrate his prior belief about the boxes' locations with the type of sampling process he observes. Note that this specific pattern of responding requires children to track simultaneously the true location of the Duck box, the Frog's belief about the location of the Duck box, and the probability of generating the observed sample from the population. The complexity of the task is necessary to distinguish children's responding to a third party's updating of his beliefs from responses children might make on other grounds (see Table 1). However, given the complexity of the task, we expected a number of children to have difficulty tracking the true location of the Duck box and the Frog's beliefs about the boxes' locations (especially as the boxes were occluded through much of the task and differed only in the relative proportion of their contents). Of course, children can only reason accurately about how the Frog might update his beliefs given the sample if they remember both the true location of the Duck box and the Frog's beliefs about the location. Thus we made an a priori decision to focus our analysis on the responses of the children who successfully answered both check questions.

2 | METHOD

2.1 | Participants and materials

Two hundred six children (mean: 66 months; range: 54–83 months) were recruited from an urban children's museum and participated in the study. The testing occurred in three waves in the following order: NL/RS_3 and NL/SS; OL/RS and OL/SS; NL/RS_5. Within each wave of testing children were randomly assigned to condition. Testing continued until 30 children passed the check questions in each condition. (See Inclusion Questions to follow.) While most of the children were white and middle class, a range of ethnicities and socioeconomic backgrounds reflecting the diversity of the Boston metropolitan area (47% European American, 24.4% African American, 8.9% Asian, 17.5% Latino, 3.9% two or more races) and the museum population (29% of museum attendees receive free or discounted admission) were represented.

Two black cardboard boxes (30 cm^3) were each separated into two sections by a cardboard barrier. The front side of both boxes was a clear plastic panel with a sheet of black felt velcroed over it. Each box had a hand-sized hole in the top. For one box, referred to as the 'Duck box', the front section was filled with 45 rubber ducks and 15 ping-pong balls. For the other box, referred to as the 'Ball box', the front section was filled with 45 ping-pong balls and 15 rubber ducks. (3:1 ratios were chosen because they are easily discriminable by preschoolers and because three consecutive ducks are far more likely to be randomly sampled from the Duck box than the Ball box.) The back sections of both boxes also contained rubber ducks and ping-pong balls, and were hidden from view. Each box was placed on a colored mat. A Frog puppet served as the agent.

2.2 | Design and procedure

We crossed the two locations where the Duck box could be at the end of the study (Old and New) and two kinds of sampling processes, sampling three ducks, from the Duck box (Random and Selective), yielding four conditions: the Old Location/Random Sampling (OL/RS) condition, the New Location/Random Sampling (NL/RS_3 ducks) condition, the Old Location/Selective Sampling condition (OL/SS), and the New Location/Selective Sampling (NL/SS) condition. We also ran a condition in the New Location/Random Sampling case in which children saw a sample of five ducks drawn from the New Location (NL/RS_5 ducks). We included this condition to ask whether children drew graded inferences that depended on the amount of randomly generated data the Frog observes (i.e., children should be more likely to think the Frog might change his mind given more randomly sampled data inconsistent with his prior beliefs).

2.2.1 | Preference phase

In all conditions, the experimenter showed the child the Duck and Ball boxes side-by-side on a table (L/R counterbalanced across participants). Each box was placed on a different colored mat, red or blue, to help children track the identities of the boxes. Initially, the felt hid the boxes' front sections. Children were given a duck and a ball, not drawn from either box to hold briefly. The experimenter then lifted the felt, revealing the front sections of both boxes and said, 'One box has mostly ducks, and one box has mostly balls. Which box has mostly ducks? Which box has mostly balls?' If the child answered incorrectly, the experimenter told the child the correct answer and repeated the questions.

Next, the experimenter introduced the agent, 'Froggy', saying, 'This is my friend Froggy!' The experimenter said, 'Froggy likes ducks better than balls.' The experimenter then asked the Frog if he wanted to play with the ball. The Frog replied, 'No, I only like ducks!' The child was asked to hand the Frog his favorite toy. The Frog's preference for ducks was established to help children track the Frog's goal of locating the Duck box. Next, both the child and the Frog learned that the boxes could move in two ways. The experimenter said, 'The boxes can move so that they are in the same place', and 'The boxes can move so that they are in different places.' (For the former, the experimenter rocked the boxes back and forth three times. For the latter, the experimenter moved the Duck box from the red mat to the blue one and the Ball box from the blue mat to the red one, or vice versa; counterbalanced across participants.) The experimenter then asked the Frog, 'Which box do you like best?' The Frog approached the Duck box and said, 'I like this box! I like the Duck box!' The experimenter returned the boxes to their original locations. The experimenter asked the child to point to the box the Frog preferred; all children answered correctly.

2.2.2 | Belief phase

Next, the experimenter told the child that the Frog was tired and hid him under the table. Children watched as the experimenter re-covered the front of both boxes with the felt. For children in the Old Location conditions, the experimenter rocked the boxes back and forth saying, 'I'm going to move the boxes so that they are in the same place.' For children in the New Location conditions, the experimenter switched the locations of the boxes saying, 'I'm going to play a trick on Froggy! I'm going to move the boxes so that they are in different places.'

Inclusion questions

In both conditions, the experimenter then asked children two questions to check that they understood the true locations of the boxes (*location check*) and the Frog's beliefs about the boxes (*belief check*). The *location check* question was, 'Where is the Duck box?' The *belief check* question was, 'Where does Froggy think the Duck box is?'

2.2.3 | Sampling phase

The experimenter brought the Frog back saying, 'Look, Froggy is back!' The experimenter asked the Frog to watch the two boxes and then responded to a pretend phone call saying, 'Hello? Oh, you want me to take three (five in the NL/RS_5 condition) ducks from the box on the red (blue) mat?' (The experimenter always named the actual location of the Duck box.) We included the phone call to dispel any impression that the experimenter was pedagogically sampling from the box in order to teach the Frog (or child) the actual location of the Duck box. Note that pedagogical sampling is always selective, but intentional sampling can be either random or selective: one can intentionally pull objects out at random or intentionally choose particular objects (see e.g., Gweon et al., 2010, for discussion). In the Random Sampling conditions, the experimenter looked over her shoulder (i.e., not into the box) and reached through the hole into the Duck box three times in rapid succession, drawing out a duck each time and counting out 'One, two, three (four, five, only in the NL/RS_5 condition)'. In the Selective Sampling conditions, the experimenter peered through the hole into the Duck box and kept her hand inside the box for approximately 2 seconds before retrieving a duck. She counted, 'One… two… three' after finding each duck. After sampling three ducks from the box and ending the pretend phone call, the experimenter asked, 'Froggy, did you see that?' to which the Frog replied, 'Yes.'

2.2.4 | Test phase

In the final phase of the experiment after the sample of three ducks was drawn, children were asked the critical test question: 'Where does Froggy think the Duck box is now?'

3 | RESULTS

3.1 | Inclusion questions

Children's responses were coded from videotape by the first authors. Forty-seven percent of the data was coded by a second coder, blind to condition and hypotheses. Inter-coder reliability was high (Kappa = .95, 98% agreement).

We coded children's responses to the location ('Where is the duck box?') and belief ('Where does Froggy think the duck box is?') check questions. Of the 206 children tested, 73% (N = 150) answered both questions correctly ('trackers') and 27% (N = 56) answered one or both of the check questions incorrectly ('non-trackers'). The number of children excluded for failing only the location question, only the belief question or both by condition is as follows: NL/RS_3: location: 6; belief: 2; both: 4; NL/RS_5: location: 0; belief: 7; both: 8; NL/SS: location: 1; belief: 12; both: 4; OL/RS: location: 4; belief: 3; both: 0; OL/SS: location: 1; belief: 4; both: 0.

Non-trackers were younger than trackers (non-trackers: M = 64 months; trackers: M = 67 months; $t(204) = 2.52$, $p = .01$, $d = .39$). Non-trackers may have subsequently given responses about the Frog's belief that did not reflect information necessary to make accurate rational inferences on behalf of the Frog so we excluded these children from our primary analysis. This resulted in a final sample of N = 150 (53% female[1]) children across the five conditions: NL/RS_3 (n = 30, m_{age} = 65 mo.; *range*: 54–78 months), NL/SS (n = 30, m_{age} = 66 mo.; *range*: 54–82 months), OL/RS (n = 30, m_{age} = 66 mo.; *range*: 54–81 months), OL/SS (n = 30, m_{age} = 68 mo.; *range*: 55–82 months), and NL/RS_5 (n = 30, m_{age} = 69 mo.; *range*: 57–83 months). Age in months did not differ across conditions ($F(4, 145) = 1.05$, $p = .38$).

Note that more children failed the inclusion questions in the New Location condition than the Old Location condition (unsurprisingly since the New Location condition involved tracking both a change of location and representing a false belief). On average, 8.67 more children were excluded for failure to track the boxes' locations and/or the Frog's beliefs in the three New Location conditions than the two Old Location conditions (a 33% exclusion rate versus a 17% exclusion rate;

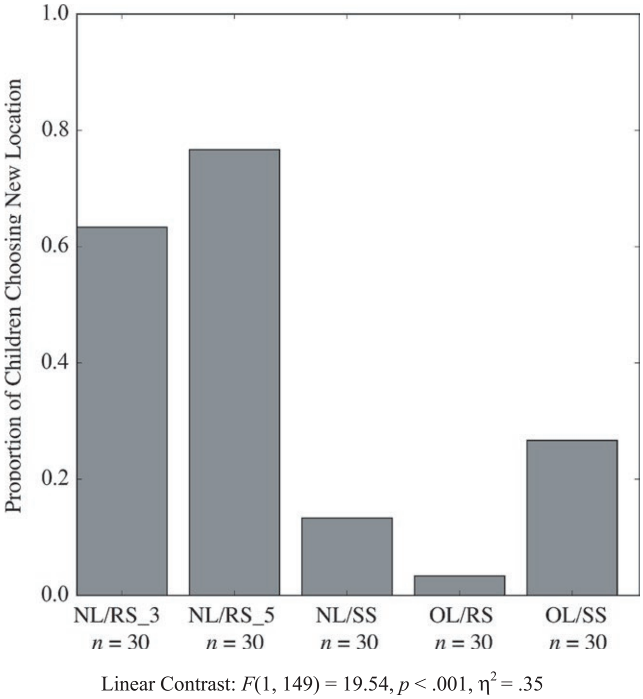


FIGURE 2 Proportion of children who passed the inclusion criteria ('trackers') who chose the New location in each condition in response to the test question about the Frog's belief

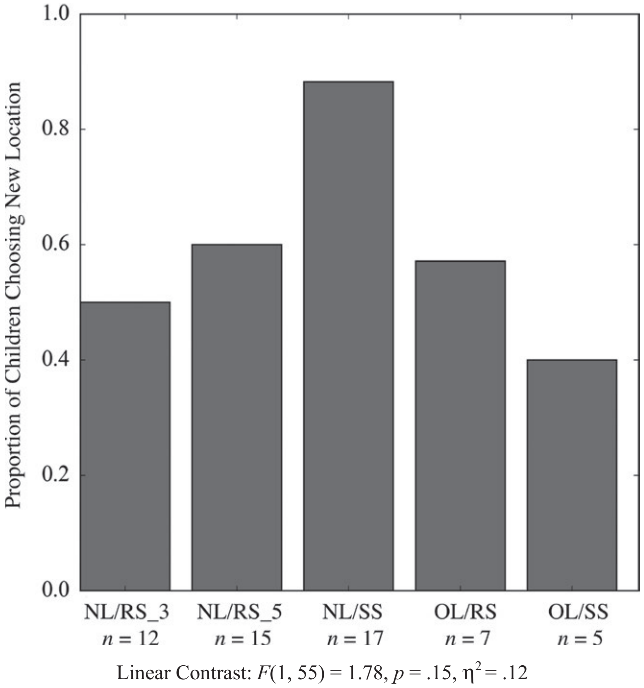


FIGURE 3 Proportion of children who failed the inclusion criteria ('non-trackers') who chose the New Location in each condition in response to the test question about the Frog's belief

$p = .01$). This raises the possibility that the included sample of children in the New Location might differ from those in the Old Location condition in any of a number of ways (e.g., including being more attentive or motivated, having better theory of mind or executive function skills, or differing with respect to other cognitive abilities).

Critically, however, the rational learning account does not predict better, or even simply uniformly *different* performance in the New Location conditions than the Old Location conditions (predictions whose investigation could be confounded to the degree that one group of children met more stringent inclusion criteria than the other). Rather, it predicts a precise pattern of responses depending jointly on the Frog's initial beliefs about the boxes' location, the sampling process, and the amount of evidence observed. That is, this account makes predictions within each condition (where there are no differences in exclusion rates) and also predicts both commonalities and differences across conditions. Neither the prediction that, within each condition, children should be more likely to expect the Frog's beliefs to be informed by randomly than selectively sampled evidence, nor the prediction that children should draw stronger inferences for the Old than the New Location condition given randomly (but not selectively) generated evidence, can be accounted for by an overall difference between the two conditions.

3.2 | Test question

Because we had a priori hypotheses about the pattern of results, we performed planned linear contrasts. We formalized the prediction that the responses in the New Location/Random Sampling conditions would differ from the other three conditions, and that the other three conditions would not differ from each other by conducting the analyses with following weights: New Location/Random Sampling with three ducks (3), the New Location/Selective Sampling (−2), the Old Location/ Random Sampling (−2), the Old Location/Selective Sampling (−2), and the New Location/Random sampling with five ducks (3).

For the 150 children who recalled the Frog's belief as well as the boxes' actual locations, the linear contrast was significant ($F(1, 149) = 19.54$, $p < .001$, $\eta^2 = .35$). Children were significantly more likely to believe that the Frog had updated his belief to the New Location in the New Location/Random Sampling conditions than in the other conditions (percentage of children choosing New Location by condition: NL/RS_3: 63%; NL/RS_5: 77%; NL/SS: 13%; OL/RS: 3%; OL/SS: 27%; see Figure 2).

By contrast, for the children who answered at least one of the check questions incorrectly (the non-trackers), the linear contrast was not significant ($F(1, 55) = 1.78$, $p = .15$, $\eta^2 = .12$). Instead, children appeared to either respond at chance or respond to the last location where they had seen the ducks (see Figure 3). Crucially, these results suggest that the children who met the inclusion criteria were not simply defaulting to some baseline response pattern but were instead responding as predicted: inferring that the Frog would rationally update his beliefs from the data.

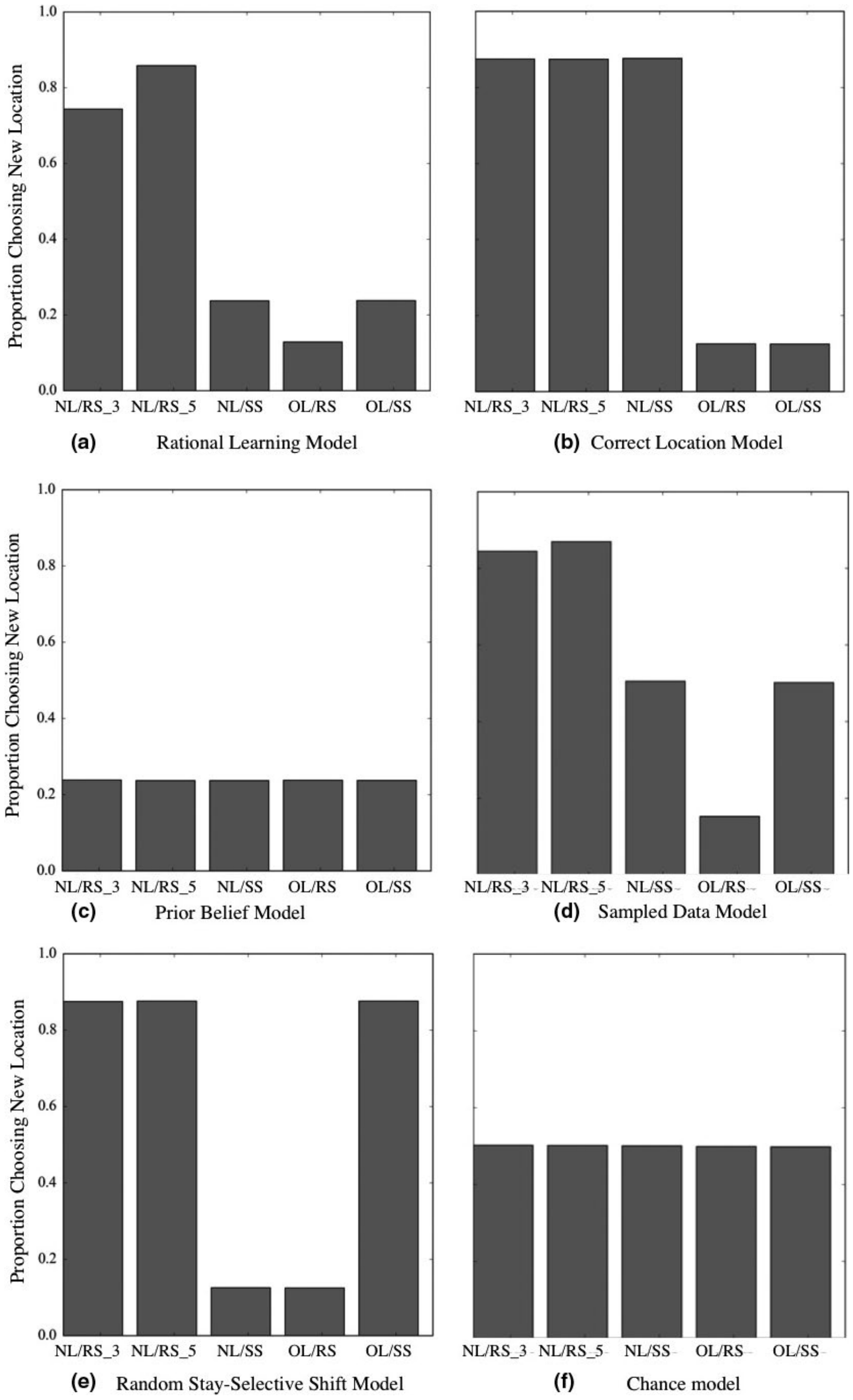
We restrict our analyses to children who pass the inclusion criteria because there is no clear way to interpret the responses of children who lost track of the boxes' location or failed to represent the Frog's initial beliefs. However, the linear contrast remains significant if all 206

TABLE 2 Bayes factor analyses comparing the Rational Learning model with the alternative models

	Correct Location	Prior Belief	Random-Stay/Selective-Shift	Sampled Data	Chance
Rational Learning:	33.73: 1	42.98: 1	26.32: 1	45.80: 1	146.20: 1

children are included ($F(1, 205) = 10.314$, $p < .001$, $\eta^2 = .22$), suggesting that the results are robust to the exclusion criteria.

Looking within each condition, children chose the Old Location significantly more often than chance in all conditions (percentage of children choosing Old Location: NL/SS: 87%, $p < .001$; OL/RS: 97%, $p < .001$; OL/SS: 73%, $p = .02$; by binomial test) except the NL/RS_5 condition where they chose the New Location above chance (77% of children choosing New; $p = .005$ by binomial test) and the NL/RS_3 condition where they chose at chance (63% of children choosing New; $p = .20$ by binomial test).



(a) Rational Learning Model

(b) Correct Location Model

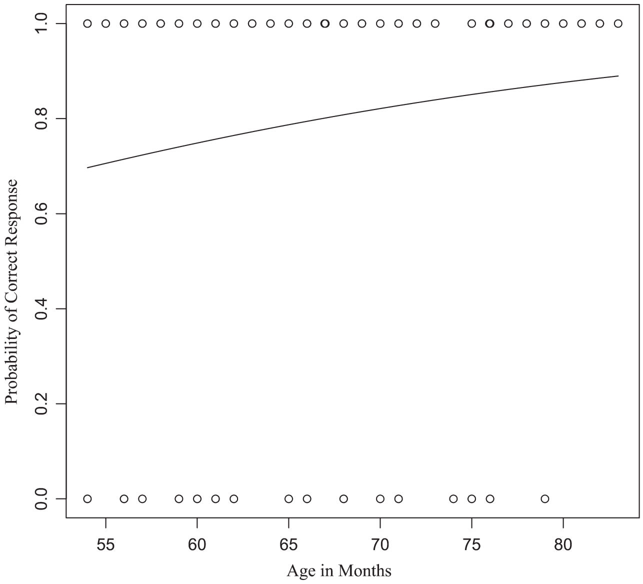
(c) Prior Belief Model

(d) Sampled Data Model

(e) Random Stay-Selective Shift Model

(f) Chance model

FIGURE 4 Predictions made by the Rational Learning Model for the rational inference model along with the five alternative models (b–f). The Rational Learning Model (a) provides the best fit to the children's responses. (See Figure 2 and Table 1.)



Logistic Regression: $\beta = 0.043(.024)$, $z = 1.791$, $p = .073$

FIGURE 5 Children's responses were coded as 1 if they were consistent with the expectation of rational learning and 0 otherwise. There was a non-significant trend for children's performance to improve with age

Our hypothesis made several key predictions about differences between conditions. First, if children expect the Frog to be sensitive to the distinction between randomly sampled and selectively sampled evidence, then given the same prior beliefs and evidence, they should expect the Frog to draw stronger inferences from randomly sampled evidence than selectively sampled evidence. Children's inferences did indeed depend on the type of evidence sampled. In the comparison between the New Location/Random Sampling_3 condition and the New Location/Selective Sampling condition children were more likely to update the Frog's false belief and infer that the Duck box was in the New Location in the Random Sampling than the Selective Sampling condition, as warranted (Fisher's exact, $p < .\ 001$). Similarly, children were more likely to think the Frog would infer that the Duck box was in the Old Location in the Old Location/Random Sampling condition compared to the Old Location/Selective Sampling condition (Fisher's exact, $p = .03$). The fact that children made comparable inferences in both conditions suggests that the results cannot be explained by differences in children's belief understanding in the two conditions (i.e., as a byproduct of the different inclusion rates in the New and Old Location conditions). Rather, children's tendency to expect the Frog's beliefs to be more influenced by randomly sampled than selectively sampled evidence in both conditions is consistent with the Rational Learning account since, indeed, randomly sampled evidence is more informative than selectively sampled evidence about the population from which it is drawn.

Also as predicted, numerically more children said the Frog would update his belief when five ducks were randomly sampled than when three ducks were randomly sampled. The difference between the NL/RS_3 condition and NL/RS_5 condition was not significant (Fisher's exact, $p = .40$); however, the graded nature of children's inferences was consistent with the predictions of the rational inference model.[2]

Finally, as predicted, children were sensitive to the Frog's prior beliefs. Given identical evidence and sampling processes, children drew different inferences when the data were sampled from the Old Location and the New Location. Thus given three ducks randomly sampled from a location, children's inferences about what he would learn from the sample depended on the Frog's prior beliefs about the location of the Duck box. Children were confident that the Frog would believe the randomly sampled data indicating that the Duck box was in the old location (97% of children in the OL/RS chose Old) but did not make as strong an inference when the randomly sampled data suggested the Duck box was in the New Location (63% of children in the NL/RS_3 chose New; OL/RS vs. NL/RS_3, Fisher's exact, $p = .002$). The analogous comparison between the selective sampling conditions was also significant. The Rational Learning model predicts that children should choose the Old Location in both selective sampling conditions because selective sampling is uninformative about the population from which it is drawn. As predicted, children interpreted identical evidence differently depending on the Frog's prior beliefs about the location: children were more likely to choose the Old Location in the OL/SS condition (73%) than they were to choose the New Location in the NL/SS condition (13%; Fisher's exact, $p < .001$).

As a further test of the hypothesis that children's judgments on behalf of the Frog reflect an expectation of rational learning, rather than any alternative model (Table 1) we can directly compare the Rational Learning model with alternative models using a Bayes factor analyses (see Gelman et al., 2013). As is clear in Table 2 and Figure 4, the Rational Learning Model outperforms all of the alternative models in predicting the data. See Appendix S1 for details.

Finally, we looked at whether the ability to make the rational inference on behalf of the Frog changed between four-and-a-half and six years. We coded children's responses as a '1' if they responded with the Old Location in the Old Location/Random Sampling, Old Location/Selective Sampling, and New Location/Selective Sampling conditions and with the New Location in the New Location/Random Sampling conditions and a '0' if they responded otherwise. The logistic regression was marginally significant, suggesting a trend for older children to be more likely to expect others to rationally update their beliefs, $\beta = 0.043(.024)$, $z = 1.791$, $p = .073$. See Figure 5.

4 | DISCUSSION

The results of the current study suggest that young children not only expect agents to act rationally with respect to their goals (Gergely & Csibra, 2003), they expect other agents to learn rationally from data. To make inferences on behalf of another agent, children needed to integrate the agent's prior beliefs with the evidence the agent observed and the way the evidence was sampled. Children were inclined to believe that the Frog would change his mind only when there was strong evidence against the Frog's prior belief (i.e., in the New Location/Random Sampling conditions). Children did not expect the Frog to change his mind when the evidence was consistent with his prior beliefs (Old Location/Random Sampling; New Location/Selective Sampling), or when the evidence may have conflicted with the Frog's prior beliefs but was weak and thus provided little ground for belief revision (Old Location/Selective Sampling).

Although even the youngest children in our sample were able to draw inferences about how a third party would update his beliefs from data, this study provides suggestive evidence that this ability might increase with age. Future research might look both at how children's ability to draw inferences about others' learning changes over development and investigate the origins of this sensitivity earlier in childhood. A basic understanding of how evidence affects others' beliefs (e.g., the understanding that seeing leads to knowing; Onishi & Baillargeon, 2005; Pratt & Bryant 1990; Senju, Southgate, Snape, Leonard, & Csibra, 2011) emerges very early. This knowledge, together with the ability to make predictions about rational action, opens up the possibility that in simpler contexts, even younger children might be able to draw inferences about how third parties might update their beliefs from data. It is also possible that children's representations of the processes that underlie belief revision may support the emergence of broader abilities in interpretive theory of mind (Astington et al., 2002; Carey & Smith, 1993, Chandler & Carpendale, 1998; LaLonde & Chandler, 2002; Myers & Liben, 2012; Pillow & Mash, 1999; Ross et al., 2005; Ruffman et al., 1993); future research might investigate the relationship between understanding that evidence conflicts with prior knowledge and understanding that evidence can be ambiguous depending on prior knowledge.

As discussed, children might have made a wide range of other inferences. In particular, they might have assumed that the Frog's beliefs would mirror their own; they might have recognized that the Frog's beliefs depended on epistemic access to the location of the box but failed to recognize that the Frog might update his beliefs based on inferential evidence, or they might have expected the Frog to attend to the sampled evidence but not have expected the Frog to integrate this evidence with his prior beliefs. Yet, children in this study were able to make predictions about what the Frog would think about the location of the Duck box given the evidence, even though they themselves always knew the true location of the Duck box. Moreover, children were able to draw different inferences depending on the ambiguity of the evidence, showing different patterns of responding in the Random and Selective Sampling conditions. This suggests that children can draw inferences that are sensitive both to the distinction between their own and others' prior knowledge, and to the strength of the data that others observe. We believe this finding is broadly consistent with an emerging body of literature suggesting that children make relatively nuanced decisions about when and what to learn from others (Bonawitz et al., 2011; Corriveau, Fusaro, & Harris, 2009; Gweon, Pelton, Konopka, & Schulz, 2014; Jaswal, 2010; Jaswal, Croft, Setia, & Cole, 2010; Koenig, Clement, & Harris, 2004; Koenig & Harris, 2005; Stiller, Goodman, & Frank, 2015). Our study extends the literature by suggesting that children also make relatively nuanced decisions about how and when children will expect others to learn.

The current study, however, does not indicate how broadly this ability extends, nor does it suggest the conditions under which children might fail to expect others to rationally update their beliefs from data. Here we suggest an account of how children might make normative judgments on behalf of third parties; future research might test the limitations of this account. Also, as discussed, the current study was motivated in part by predictions from an ideal observer model of rational inference. The results are broadly consistent with that account. However, providing a rigorous test of the quantitative predictions of the rational inference model and alternative accounts remains an important direction for future work.

As adults, we expect other agents to be rational actors not only in terms of the paths they take towards their goals, but also in terms of how they reason about evidence. Here we find that children's developing theory of mind supports the same kinds of inferences. By 4½ years, children are able to integrate others' prior knowledge and observed evidence to support predictions about when others will retain their beliefs and when they will change their minds.

ACKNOWLEDGEMENTS

We thank the Boston Children's Museum and the families who participated in this research. We also thank Eleanor Holton and Rebecca Leibowitz for help with coding the data and members of the Early Childhood Cognition Lab for helpful comments and discussion. This material is based upon work supported by the Center for Brains, Minds, and Machines (CBMM), funded by NSF STC award CCF-1231216 and was funded by a National Science Foundation Career Award (#0744213) to LS.

NOTES

[1] Information on the children's gender was available only for 81% of the children; the reported percentage reflects this sub-sample.

[2] Note that although ages did not differ significantly across conditions, the mean age of children in the NL/RS_5 condition was 69 months, compared to 65 months for children in the NL/RS_3 condition. We are grateful to an anonymous reviewer for pointing out the possibility that this age difference may have contributed to children's stronger inferences in the NL/RS_5 condition.

REFERENCES

Astington, J.W., Pelletier, J., & Homer, B. (2002). Theory of mind and epistemological development: The relation between children's second-order false-belief understanding and their ability to reason about evidence. *New Ideas in Psychology*, *20*, 131–144.

Baillargeon, R., Scott, R.M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences*, *14*, 110–118.

Baker, C.L., Saxe, R., & Tenenbaum, J.B. (2009). Action understanding as inverse planning. *Cognition*, *113*, 329–349.

Baker, C.L., Saxe, R., & Tenenbaum, J.B. (2011). Bayesian theory of mind: Modeling joint belief-desire attribution. In *Proceedings of the Thirty-Third Annual Conference of the Cognitive Science Society* (pp. 923–928).

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

Carey, S., & Smith, C. (1993). On understanding the nature of scientific knowledge. *Educational Psychologist*, *28*, 235–251.

Carpendale, J.I., & Chandler, M.J. (1996). On the distinction between false belief understanding and subscribing to an interpretive theory of mind. *Child Development*, *67*, 1686–1706.

Chandler, M.J., & Carpendale, J.I. (1998). Inching toward a mature theory of mind. In M. Ferrari & R.J. Sternberg (Eds.), *Self-awareness: Its nature and development* (pp. 148–190). New York: Guilford Press.

Corriveau, K.H., Fusaro, M., & Harris, P.L. (2009). Going with the flow: Preschoolers prefer nondissenters as informants. *Psychological Science*, *20*, 372–377.

Csibra, G., Bíró, S., Koós, O., & Gergely, G. (2003). One-year-old infants use teleological representations of actions productively. *Cognitive Science*, *27*, 111–133.

D'Andrade, R. (1987). A folk model of the mind. In D. Holland & N. Quinn (Eds.), *Cultural models in language and thought* (pp. 112–148). Cambridge: Cambridge University Press.

Denison, S., & Xu, F. (2014). The origins of probabilistic inference in human infants. *Cognition*, *130*, 335–347.

Dennett, D. (1987). *The intentional stance*. Cambridge, MA: MIT Press.

Fodor, J. (1987). *Psychosemantics: The problem of meaning in the philosophy of mind*. Cambridge, MA: MIT Press.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., & Rubin, D.B. (2013). *Bayesian data analysis* (3rd edn). London: Chapman & Hall/CRC Press.

Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naïve theory of rational action. *Trends in Cognitive Sciences*, *7*, 287–292.

Gergely, G., Nádasdy, Z., Csibra, G., & Bíró, S. (1995). Taking the intentional stance at 12 months of age. *Cognition*, *56*, 165–193.

Gopnik, A., & Wellman, H.M. (2012). Reconstructing constructivism: Causal models, Bayesian learning mechanisms, and the theory theory. *Psychological Bulletin*, *138*, 1085–1108.

Gweon, H., Pelton, H., Konopka, J.A., & Schulz, L.E. (2014). Sins of omission: Children selectively explore when teachers are under-informative. *Cognition*, *132*, 335–341.

Gweon, H., Tenenbaum, J., & Schulz, L.E. (2010). Infants consider both the sample and the sampling process in inductive generalization. *Proceedings of the National Academy of Sciences of the United States of America*, *107*, 9066–9071.

Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.

Jara-Ettinger, E., Baker, C.L., & Tenenbaum, J.B. (2012). Learning what is where from social observations. In *Proceedings of the Thirty-Fourth Annual Conference of the Cognitive Science Society* (pp. 515–520).

Jaswal, V.K. (2010). Believing what you're told: Young children's trust in unexpected testimony about the physical world. *Cognitive Psychology*, *61*, 248–272.

Jaswal, V.K., Croft, A.C., Setia, A.R., & Cole, C.A. (2010). Young children have a specific, highly robust bias to trust testimony. *Psychological Science*, *21*, 1541–1547.

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.

Koenig, M.A., & Harris, P.L. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child Development*, *76*, 1261–1277.

Kushnir, T., Xu, F., & Wellman, H.M. (2010). Young children use statistical sampling to infer the preferences of other people. *Psychological Science*, *21*, 1134–1140.

Lalonde, C.E., & Chandler, M.J. (2002). Children's understanding of interpretation. *New Ideas in Psychology*, *20*, 163–198.

Myers, L.J., & Liben, L.S. (2012). Graphic symbols as 'the mind on paper': Links between children's interpretive theory of mind and symbol understanding. *Child Development*, *83*, 186–202.

Onishi, K.H., & Baillargeon, R. (2005). Do 15-month-old-infants understand false beliefs? *Science*, *308*, 255–258.

Pillow, B.H., & Mash, C. (1999). Young children's understanding of interpretation, expectation and direct perception as sources of false belief. *British Journal of Developmental Psychology*, *17*, 263–276.

Pratt, C., & Bryant, P. (1990). Young children understand that looking leads to knowing (so long as they are looking into a single barrel). *Child Development*, *61*, 973–982.

Ross, H.S., Recchia, H.E., & Carpendale, J.I. (2005). Making sense of divergent interpretations of conflict and developing an interpretive understanding of mind. *Journal of Cognition and Development*, *6*, 571–592.

Ruffman, T., Perner, J., Olson, D.R., & Doherty, M. (1993). Reflecting on scientific thinking: Children's understanding of the hypothesis–evidence relation. *Child Development*, *64*, 1617–1636.

Schulz, L.E. (2012). The origins of inquiry: Inductive inference and exploration in early childhood. *Trends in Cognitive Sciences*, *16*, 382–389.

Senju, A., Southgate, V., Snape, C., Leonard, M., & Csibra, G. (2011). Do 18-month-olds really attribute mental states to others? A critical test. *Psychological Science*, *22*, 878–880.

Skerry, A.E., Carey, S.E., & Spelke, E.S. (2013). First-person action experience reveals sensitivity to action efficiency in prereaching infants. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, 18728–18733.

Sodian, B., Taylor, C., Harris, P.L., & Perner, J. (1991). Early deception and the child's theory of mind: False trails and genuine markers. *Child Development*, *62*, 468–483.

Sodian, B., & Wimmer, H. (1987). Children's understanding of inference as a source of knowledge. *Child Development*, *58*, 424–433.

Stiller, A.J., Goodman, N.D., & Frank, M.C. (2015). Ad-hoc implicature in preschool children. *Language Learning and Development*, *11*, 176–190.

Tenenbaum, J.B., Kemp, C., Griffiths, T.L., & Goodman, N.D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, *331*, 1279–1285.

Wellman, H.M. (2014). *Making minds: How theory of mind develops*. Oxford: Oxford University Press.

Wellman, H.M., Cross, J.W., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, *72*, 655–684.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*, 103–128.

Xu, F., & Denison, S. (2009). Statistical inference and sensitivity to sampling in 11-month-old infants. *Cognition*, *112*, 97–104.

Xu, F., & Garcia, V. (2008). Intuitive statistics by 8-month-old infants. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 5012–5015.

Zaitchik, D. (1991). Is only seeing really believing? Sources of the true belief in the false belief task. *Cognitive Development*, *6*, 91–103.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Magid RW, Yan P, Siegel MH, Tenenbaum JB, Schulz LE. Changing minds: Children's inferences about third party belief revision. *Dev Sci*. 2017;00:e12553. https://doi.org/10.1111/desc.12553

APPENDIX

Computational Model

To help clarify our proposal and specify what counts as 'rational inference' in these contexts, we developed a computational model that provides quantitative predictions for each experimental condition. The model specifies how a rational agent would behave when presented with the same task that we gave our participants. Although many studies have used Bayesian models to assess children's ability to update their *own* prior beliefs from data (see Gopnik & Wellman, 2012; Schulz, 2012; and Tenenbaum et al., 2011, for reviews) to our knowledge, this is the first attempt to consider children's ability to predict when *another* agent will (or will not) change his mind by considering both that agent's access to the data and his prior beliefs. Finally, note that in suggesting that children's rational inferences on behalf of a third party can be captured by a Bayesian inference model, we do not mean to suggest that children have conscious, meta-cognitive access to these computations; rather, we suggest that such sophisticated computations may underlie the many implicit, rapid, accurate judgments that support everyday social cognition. Figure 4 in the Main Text displays the predictions of our model for each of the candidate hypotheses of Table 1 in the Main Text.

The model is specified at two levels. First, we built a model of the Frog as a rational learner, given the information that he has available to him. Then, we modeled children's rational inferences about the Frog. Thus two levels of rational inference are represented: the Frog's beliefs about the location of the box, and the child's beliefs about the Frog's beliefs.

We adopt a Bayesian framework for modeling both these levels of rational inference. Bayesian inference models a learning event as an interaction of two factors: the agent's prior beliefs about a hypothesis, before seeing new data: *p(h)*, and the probability that the hypothesis is true given the newly observed data, the likelihood *p(D | h)*. These combine to yield the agent's updated posterior belief *p(h | D)*. Given new data bearing on a hypothesis, Bayes' rule specifies how a rational agent should update her beliefs as:

$$p(h|D) \propto p(D|h)p(h)$$

We now turn to the model of the Frog's inference, from the perspective of an ideal observer (which we can consider the child as approximating). On each experimental trial, the experimenter draws ducks from the Duck box, either randomly or selectively, and the boxes may or may not have been switched. At that point, both the child and the Frog know whether the sample is drawn randomly or selectively but only the child knows whether the boxes have been switched. However, the Frog has some prior belief *pswitch* about whether the boxes were switched in his absence (given the demonstration that they can be switched), which is equivalent to having a prior belief about which box the ducks are being drawn from. We can specify these as *p(hduck)* = *pswitch* for the Duck box and *p(hball)* = 1 − *pswitch* for the Ball box. The Frog must integrate this prior belief with his observation of three (or five) ducks being drawn from the box. Under random sampling, the probability of drawing *n* ducks and zero balls, with replacement,[1] from the duck box is $h_{duck} = \left(\frac{45}{60}\right)^n = \left(\frac{3}{4}\right)^n$; similarly the probability of drawing *n* ducks from the ball box is $h_{ball} = \left(\frac{1}{4}\right)^n$ (these quantities specify the likelihood of the data given each hypothesis). Under selective sampling, the experimenter explicitly reached into the box to pull out a duck and thus the probability is 1 from each box. The posterior beliefs *p(hduck | D)* and *p(hball | D)* are then given by Bayes' theorem above:

$$p(h_{duck}|D) \propto p_{switch} \times \left(\frac{3}{4}\right)^n \quad \text{and} \quad p(h_{ball}|D) \propto p_{switch} \times \left(\frac{1}{4}\right) \tag{1}$$

in the case of random sampling, and

$$p(h_{duck}|D) \propto p_{switch} \quad \text{and} \quad p(h_{ball}|D) \propto p_{switch} \tag{2}$$

in the case of selective sampling.

Thus we have a posterior distribution over the two hypotheses, where the posterior probability that the sample is drawn from the Duck box increases as the number of randomly sampled ducks increases, and remains equal to the prior under selective sampling. This reflects our intuition that the evidence is stronger with each new randomly sampled duck and unchanged with each selectively sampled duck.

Having specified a rational model of the Frog's inference, we now describe our model of the experimental participants. We propose that children can approximately simulate the above inference, and when asked to say where they think the Frog thinks the duck box is, they report the output of this computation, subject to two approximations. As is standard practice when modeling behavioral responses (Denison, Bonawitz, Gopnik, & Griffiths, 2013; Gweon et al., 2010; Xu & Tenenbaum, 2007), we assume that children probability match; that is, the frequency with which they select responses is proportional to the posterior probability of each hypothesis. In a population of participants, this rule gives a distribution of responses that mimics the distribution of posterior beliefs, and it is an efficient scheme for approximating probabilistic inference (Vul, Goodman, Griffiths, & Tenenbaum, 2014). We also consider the possibility that on each trial, there is a nonzero probability that children may have been inattentive or confused. We therefore include a noisy response parameter, *perror*, estimating the probability that a participant gives a box choice selected uniformly at random, instead of the response predicted by the model. Thus our model at this point has two parameters: the Frog's prior belief, *p_switch*, about whether the boxes were switched, and the noisy response parameter, *p_error*. To estimate *pswitch*, we used the ratio of children's responses on the initial belief question:

$$p_{switch} = \frac{28}{178} = .157$$

We have no analogous way to derive a plausible independent and numerically precise estimate of *perror*. For the results displayed in Figure 4, we set *perror* = .25; as children had to pass two inclusion checks, at most 25% of included children could have been answering at chance.

While we have described our model mathematically, it is possible to implement this model implicitly by simple sampling operations, without making any explicit statistical calculations. We describe one such implementation written in the probabilistic programming language Church (Goodman, Mansinghka, Roy, Bonawitz, & Tenenbaum, 2008; Goodman & Tenenbaum, 2014).

We implemented the Rational Learning model, and all of the alternative models presented in Table 1 in the probabilistic programming language Church (Goodman et al., 2008; Goodman & Tenenbaum, 2014). We used the webchurch implementation, available at https://github.com/probmods/webchurch or interactively at https://probmods.org/play-space.html. To evaluate the following Church code, copy and paste the code text into the environment available in the latter link.

The following code block is sufficient to reproduce all of the model predictions described here; to obtain the predictions for individual conditions given a specified model, modify the variables num-draws, actual-switch, and sampling-manner as described in the text. To obtain the predictions of different (alternative) models, the predictions of the different (alternative) models (a) - (f), modify the value of the variable which-model to the appropriate 'a-'f in the code listed in Data S1.

APPENDIX NOTE

[1] While the experiment used sampling without replacement, our model used sampling with replacement because the analysis is conceptually simpler and for large populations (i.e., the 60 objects in the box here) the difference between the distributions underlying sampling with and without replacement is negligible.

APPENDIX REFERENCES

Denison, S., Bonawitz, E., Gopnik, A., & Griffiths, T.L. (2013). Rational variability in children's causal inferences: The sampling hypothesis. *Cognition*, *126*, 285–300.

Goodman, N., Mansinghka, V., Roy, D.M., Bonawitz, K., & Tenenbaum, J. (2008). Church: A language for generative models with non-parametric memoization and approximate inference. In *Proceedings of Uncertainty in Artificial Intelligence*.

Goodman, N.D., & Tenenbaum, J.B. (electronic). Probabilistic models of cognition. Retrieved from http://probmods.org.

Vul, E., Goodman, N., Griffiths, T.L., & Tenenbaum, J.B. (2014). One and done? Optimal decisions from very few samples. *Cognitive Science*, *38*, 599–637.

Xu, F., & Tenenbaum, J.B. (2007). Word learning as Bayesian inference. *Psychological Review*, *114*, 245–272.