Young Children's Understanding of Prior and Posterior Probability

Stephanie Alderete, Ellen Tong, Wenqing Anna Cao, Steven Piantadosi & Fei Xu
Department of Psychology, University of California, Berkeley

Abstract

This study investigates 4-6-year-olds' ability to reason about prior and posterior probabilities, and how they update their decisions based on new evidence. Across two experiments, children made a prior probability guess and then, after receiving additional information, a posterior probability guess. Our findings suggest that children as young as four can make accurate prior probability guesses and in some cases, update them when given new evidence. Children's ability for probability updating improves with age. These results suggest that the ability to reason about posterior probabilities emerges earlier than previously thought, by age 4.

Keywords: Probabilistic Reasoning; Probability Updating; Cognitive Development

Introduction

Decision-making under uncertainty often requires integrating prior knowledge with new evidence. Posterior probabilities - the ability to integrate prior probabilities with new evidence – is fundamental to effective decision-making. Despite its importance in cognition and development, limited research has examined children's understanding of posterior probabilities.

A number of studies provide evidence that infants, toddlers, and 3- to 5-year-old children are capable of probabilistic reasoning (e.g., Denison & Xu, 2010, 2014; Kushnir et al. 2010; Ma & Xu, 2011; Schulz et al. 2007; Téglás et al. 2011; Wellman et al. 2016; Xu & Garcia, 2008; Yeung et al. 2016; see Denison & Xu, 2019 for review). Yet other studies document preschoolers' difficulty in making probabilistic judgments (Girotto & Gonzalez, 2008; Girotto et al., 2016; Placì, Fischer, & Rakoczy, 2020; Téglás et al., 2007). It remains an open question at what age children show a robust success in probabilistic reasoning.

Girotto and Gonzalez (2008) found that children do not make accurate prior or posterior probability judgments until at least age 5. In their study, Italian schoolchildren completed a betting task with two characters (Black and White) and colored chips (e.g., four black circles, one black square, three white squares). Children made a prior probability guess about a randomly drawn chip's owner, then a posterior guess after learning its shape. For half of the children, the shape matched the more probable color from prior trials (posterior confirmation), for the others, the shape suggested a different probable color (posterior disconfirmation). Children 5 and older succeeded in both judgments, while 4-year-olds failed at both. Fontanari and colleagues (2014) replicated these findings with older children lacking formal schooling

One methodological difference that may help explain some of the mixed findings is the ratios used in the experiments. In particular, in the study by Girotto and Gonzalez (2008), children were sometimes given an array of chips consisting of four black circles, one black square, and three white squares – a ratio of 5:3 for black vs. white chips and a ratio of 3:1 if we only consider the square chips. Most other studies have used 3:1, 4:1 or even bigger ratios (e.g., Kushnir et al. 2010; Téglás et al. 2007; Xu & Garcia, 2008; etc.). It might be the case that the 5:3 ratio presented in the prior probability trials is not sufficient for younger children to use reliably in their judgments. Furthermore, their study asked the children which character (Mr. Black or Mr. White) the randomly drawn chip (black or white) belongs to, an indirect assessment that may be more difficult for younger children. Lastly, some previous research has shown cultural differences in probabilistic reasoning (e.g., Italian vs. Chinese elementary school children in a Bayesian reasoning task, Zhu & Gigerenzer, 2006; Pighin, Girotto, & Tentori, 2017), highlighting the need to examine it cross-culturally. These considerations motivated the current study.

The main goal of the current study is to investigate 4- to 6-year-old U.S. children's ability to reason about prior and posterior probabilities, with a study design modified from Girotto and Gonzalez (2008). Specifically, we refine the original task by (1) removing the character-based reasoning so children state the more probable color rather than choosing between the two characters (2) increasing the number of chips from 5:3 (a small ratio) to 12:4 (a larger ratio, 3:1) to improve ratio discrimination in the prior trials, (3) increasing the ratio from 3:1 to 1 to improve ratio discrimination in the posterior confirmation trials, and (4) using a within-subjects design with multiple trials.

In our study, children guessed the color of a block randomly drawn from a container by an agent before and after receiving shape information. At the start of each trial, children saw colored blocks in a box (e.g., an array of 3 black squares, 1 blue square, 8 blue stars, Figure 1) placed in an opaque container and shaken. The agent selected a block, but its identity was hidden. Children then made a prior probability guess about the block's color, with the most likely color (e.g., blue) being the correct response.

After the prior probability trial, children made a posterior guess about the block's color based on its shape. The shape either confirmed (posterior confirmation) or contradicted (posterior disconfirmation) their initial guess. In posterior confirmation trials, the shape matched the more probable color (e.g., a star, so the child should keep guessing blue). In posterior disconfirmation trials, the shape suggested a less probable color (e.g., a square, so the child should switch to black). If children update probabilities, they should adjust their guess accordingly.

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

Experiment 1

Methods. Experiment 1 was approved by the [Redacted] Institutional Review Board, and pre-registered on as.predicted.org (https://aspredicted.org/2mzn-y8mr.pdf).

Participants. The final sample included 48 5- and 6-year-old children (mean age = 6.01, range = 5.07–6.97, SD = 0.52; 21 female). The sample size was determined based on prior studies investigating posterior probability in young children (Girotto & Gonzalez, 2008). Three additional children were excluded from the study (one child failed the color training trials and two children failed the shape training trials; see below). Participants were tested in person at a local science museum. Consent was obtained prior to the study, and participants received a small prize as compensation.

Materials. The materials included animated images of a robot, an opaque jar, and a box with different colored and shaped blocks, created using Microsoft PowerPoint. All stimuli were presented on a laptop. Data were analyzed in R using packages such as Tidy Verse, lme4, rstatix, and dplyr.

Design. Experiment 1 used a within-subject design. Children were trained before the test trials to ensure that they could identify colors and shapes, and answer correctly the majority color or shape of blocks in a box.

Each participant completed eight test trials, divided into four sets. Each set of test trials included one prior probability trial (henceforth *prior trials*) and one posterior probability trial (henceforth *posterior trials*). That is, children made a prior probability guess followed by a posterior probability guess. There were two types of posterior probability trials: (1) the more probable color in the posterior trial was the same as that in the prior trial (henceforth *posterior confirmation trials*) and (2) the more probable color in the posterior trial was different from that in the prior trial (henceforth *posterior disconfirmation trials*).

Children were assigned to one of four counterbalanced orders, which controlled for the following features: (1) the order in which the shape of the blocks in the array was introduced in the test trials (either left-to-right or right-to-left) and (2) the order in which the different types of test trial sets were presented (either [confirmation, disconfirmation, disconfirmation, confirmation] or [disconfirmation, confirmation, confirmation, disconfirmation]).

Procedure. Children were seated in front of a computer screen, next to the experimenter. They were asked to play a fun game where they would guess the color of the block Earl the robot was holding. Before the study, children were asked to identify the colors and shapes of different blocks. Children were excluded from the sample if they could not identify the correct color or shape of each block.

Shape Training Trials. The purpose of the shape training trials was to ensure that children could accurately say which shape of block there were more of in a box. Children completed two shape training trials. In the first trial, children were presented with a box containing four blocks: 1 blue triangle and 3 pink circles. The experimenter said to the child, “Look, here we have a box with some pink blocks and some blue blocks. Some of the blocks are circles, and some are triangles. Now, are there more circles or triangles in the box?” As the experimenter described the blocks, the corresponding block colors or shapes flashed on the screen. Children responded verbally. If the child identified the correct shape, the experimenter said, “That’s right, there are more circles than triangles.” If the child gave an incorrect answer, the experimenter corrected them by saying, “Not quite. See, there are more circles than triangles.”

The second trial was analogous to the first, except this time there were more triangles than circles in the box. Children were shown a box with 3 blue triangles and 1 pink circle and were again asked whether there were more circles or triangles in the box. Children were excluded from the final sample if they answered both trials incorrectly.

Color Training Trials. The color training trials ensured that children could identify the dominant color in a box. They completed two trials analogous to the shape training but focused on color. Children who answered both incorrectly were excluded from the final sample

Test Trials. Each child completed eight test trials, divided into four sets of prior probability and posterior probability trials. The color and shape of the blocks varied across the trials. Before the test trials began, children were told they were going to play a game to figure out the colors of different blocks that Earl the robot grabbed from a jar. They were also informed that the more questions they answered correctly, the more gold stars they would collect, which would be shown at the end of the game.

Prior trials.

In each set of trials, children made a prior probability guess about the color of the block Earl the robot had sampled from the jar. The goal was to assess whether children could accurately predict the block’s color based on the proportion of different colored blocks. During each trial, children saw Earl, the robot, standing next to an opaque jar. Below him was a box filled with different colored and shaped blocks (e.g., three black squares, one blue square, and eight blue stars) (Figure 1). The experimenter explained, “Look, here is a box with some [blue] and [black] blocks. Some of the blocks are [stars], and some are [squares].” As the experimenter described the blocks, the corresponding block colors or shapes flashed on the screen. The experimenter then said, “Now, the blocks are going into this jar [*animated blocks enter the jar*]. You can remember the blocks because they are copied on this card here [*box disappears, and a card with an image of all the blocks appears*].”

The child was then told that the blocks in the jar were going to get mixed up and saw an animation of the jar shaking. Then the experimenter said, “Now that the blocks are mixed up, Earl is going to grab one block without looking [*animation played of Earl selecting a block*]. Look, Earl is holding a block, but we can’t see which one. Looking at the card here [*gesturing to the card*], do you think Earl’s block is [black] or [blue]?”

Children responded verbally. No feedback was provided. Once the child responded, the experimenter said, “All right, thank you!” and proceeded to the posterior probability trial.

Posterior trials.

Immediately following a prior trial, children completed a posterior trial (either posterior confirmation or posterior disconfirmation) where they made a second guess about the block’s color after being told its shape.

The experimenter said to the child, “Listen closely, Earl is going to give us a hint about the block he is holding. Earl felt the block and says it’s a [e.g., **star** in posterior confirmation trials or **square** in posterior disconfirmation trials]. Now that we know Earl’s block is a [star or square], looking at the card here [*pointing to the card*], do you think Earl’s block is blue or black?”

Children responded verbally. No feedback was provided. Once the child made their selection, the experimenter said, “All right, thank you!” and moved on to the next trial.

Half of the trials were posterior confirmation trials, where the block’s shape coincided with the child’s initial guess about its color (e.g., Earl says the block is a star, so the child should continue to guess blue). The other half were posterior disconfirmation trials, where the block’s shape contradicted the child’s initial guess (e.g., Earl says the block is a square, so the child should change their answer from blue to black).

At the end of the eight trials, children were shown eight gold stars and told they did a great job.

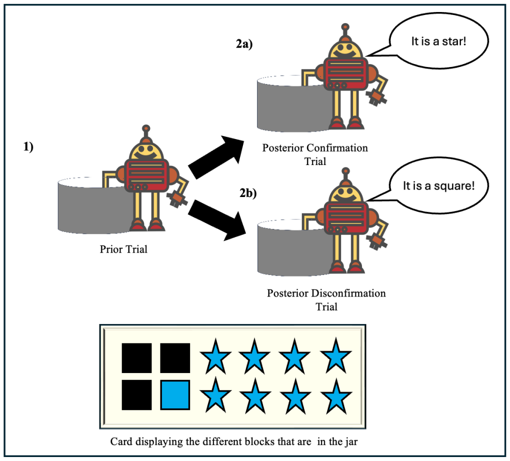

Figure 1: Test Trials: 1) prior probability trials, 2a) posterior confirmation trials, 2b) posterior disconfirmation trials

Results

The data analysis was divided into three parts. First, we conducted a preliminary analysis to assess the effects of gender, counterbalancing orders, trial number, and age on participants’ performance. Second, we analyzed the participants’ performance for the three types of test trials: prior probability trials, posterior confirmation trials, and posterior disconfirmation trials. Finally, to assess children's ability to update probabilities, we examined how often they did so accurately or got both prior and posterior trials correct. We then analyzed posterior trial performance in children who answered at least three of four prior trials correctly, ensuring the sample included those capable of probabilistic inference.

Preliminary Analysis. The preliminary analysis examined the main effects of order (Orders A–H, as described in the design section of the Methods), gender (male vs. female), age (centered and treated as a continuous variable), and trial number (Trials 1–8). A Generalized Linear Mixed Effects Model (GLMM) was fitted to predict the probability of a correct response from the fixed effects of order, gender, age, and trial number, with participant ID included as a random effect.

There was no significant main effect of order (all $p >$ .10), gender ($\hat{\beta} = 0.35$, SE = 0.27, z = 1.32, p = 0.19), trial number ($\hat{\beta}= 0.02$, SE = 0.26, z = 0.33, p = 0.74), or age ($\hat{\beta}= 0.08$, SE = 0.27, z = 0.77, p = 0.44). Thus order, trial number, gender, and age did not significantly influence participants’ responses.

Test Trials Analysis. Chance was set at 50%, as participants chose between two different colors on each test trial. Wilcoxon signed-rank tests were used to compare participants’ performance against chance.

For 5-and 6-year-olds, the average accuracy on the prior trials was 76%, significantly above chance (V = 691, p < 0.001, r = 0.81). For 5-year-olds (n= 25), the accuracy rate was 74% (SD = 0.18), significantly above chance (V = 171, p < 0.001, r = 0.82). For 6-year-olds (n = 23) accuracy rate was 77% (SD = 0.21), also above chance (V = 184, p < 0.001, r = 0.79; see Table 1).

For 5-and 6-year-olds, the average accuracy on the posterior confirmation trials was 96%, significantly above chance (V = 990, p < 0.001, r = 0.96). For 5-year-olds, the accuracy rate was 94% (SD = 0.17), significantly above chance (V = 253, p < 0.001, r = 0.94). For 6-year-olds, the accuracy rate was 97% (SD = 0.10), also significantly above chance (V = 253, p < 0.001, r = 0.98).

Finally, for 5-and 6-year-olds the average accuracy on the posterior disconfirmation trials was 71% (SD = 0.38), significantly above chance (V = 518, p < 0.001, r = 0.48). For 5-year-olds, the accuracy rate was 70% (SD = 0.40), significantly above chance (V = 157, p = 0.01, r = 0.45). For 6-year-olds, the accuracy rate was 71% (SD = 0.36), also above chance (V = 110, p < 0.001, r = 0.52).

Table 1: Performance on Trial Type by Age Group

Age Group (Years)	Prior Trials	Posterior Confirmation Trials	Posterior Disconfirmation Trials
4 (n= 24)	.71**	.96**	.50
5 (n = 25)	.74**	.94**	.70**
6 (n = 23)	.77**	.97**	.71**

Note. ** *p* < .01 indicates a significant difference from chance (.50)

Probability Updating Analysis. In the probability updating analysis, we examined how often participants answered both a prior probability trial and its corresponding posterior probability trial correctly to assess their ability to update the belief about the posterior probability. Chance was set at 25%, as each trial pair had four possible response patterns: (1) the participant correctly answered both the prior probability trial and posterior probability trial, (2) the participant correctly answered the prior probability trial but not the posterior probability trial, (3) the participant incorrectly answered the prior probability trial but the posterior probability trial correctly, or (4) the participant answered both the prior probability trial and posterior probability incorrectly.

Wilcoxon sign-ranked tests were used to compare participants' responses in the trial pairs (each consisting of a prior probability trial and its corresponding posterior probability trial) to chance (25%). Across all four sets of trials, participants answered both trial pairs correctly 61% (SD = 0.22) of the time, significantly above chance (V = 978.5, p < 0.001, r = 0.86).

When examining trial pairs that included a posterior confirmation trial, the accuracy rate was 70% (SD = 0.46), significantly above chance (V = 1135.5, p < 0.001, r = 0.84). In trial pairs that included a posterior disconfirmation trial, the accuracy rate was 58% (SD = 0.50), which was also significantly above chance (V = 983.5, p < 0.001, r = 0.61). These results indicate that participants successfully updated probabilities in both posterior confirmation trials and posterior disconfirmation trials.

Finally, we examined the posterior probability trial performance among participants who scored at least 75% on the prior trials (i.e., correctly answered three or more of the four prior trials). Of the 48 participants, 36 (75%) met this criterion. Wilcoxon signed-rank tests were used to compare their performance against chance (50%).

For these 36 participants, accuracy on the prior probability trials was 85% (SD = 0.12), significantly above chance (V = 666, p < 0.001, r = 0.90). In the posterior confirmation trials accuracy was 97% (SD = 0.16), significantly above chance (V = 595, p < 0.001, r = 0.97). In the posterior disconfirmation trials, accuracy was 68% (SD = 0.39), also significantly above chance (V = 280, p = .006, r = 0.41).

Looking at just 5-year-olds who met the criterion (n =18), accuracy on the prior probability trials was 83% (SD = 0.12), significantly above chance (V = 171, p < 0.001, r = 0.91). In the posterior confirmation trials, accuracy was 94% (SD = 0.16), significantly above chance (V = 136, p < 0.001, r = 0.95). In the posterior disconfirmation trials, accuracy was 66% (SD = 0.42), marginally significantly different from chance (V = 75, p = 0.06, r = 0.38).

Looking at just 6-year-olds who met the criterion (N =18), accuracy on the prior probability trials was 86% (SD = 0.13), significantly above chance (V = 171, p < 0.001, r = 0.91). In the posterior confirmation trials accuracy was 100% (SD = 0), significantly above chance (V = 171, p < 0.001, r = 1). In the posterior disconfirmation trials, accuracy was 69% (SD = 0.39), also significantly above chance (V = 70, p = 0.03, r = 0.46).

These findings suggest that participants who performed well on the prior probability trials were able to update probabilities effectively in both posterior confirmation trials and posterior disconfirmation trials.

Discussion

The results from Experiment 1 support the hypothesis that 5- to 6-year-old children can reason about posterior probabilities. In the prior probability trials, when given no information about the block's shape, children accurately guessed that the block was the more probable color. In the posterior probability trials, children successfully updated their guesses about the block's color when provided new information about its shape.

In the posterior confirmation trials, when the block's shape aligned with the more probable color, children continued to guess the more probable color. In the posterior disconfirmation trials, when the block's shape suggested the less probable color, children adjusted their answers accordingly. Children's success in both the prior probability trials and the posterior probability trials suggests that they can make probabilistic guesses and update them based on new information.

Experiment 2

Experiment 2 aimed to determine whether younger children can also reason about posterior probabilities.

Methods. Experiment 2 was approved by the [Redacted] Institutional Review Board, and pre-registered on as.predicted.org (https://aspredicted.org/gd9q-833k.pdf).

Participants. The final sample comprised 24 4-year-old children (mean age = 4.45, range = 4.02–4.99, SD = 0.27; 18 female). Three additional children were excluded from the study (two failed the shape training trials and one failed the color training trials). Participants were tested in person at a local science museum. Consent was obtained prior to the study, and participants received a small prize as compensation.

Materials, Design, and Procedure. The materials, design, and procedure were the same as in Experiment 1.

Results

Experiment 2 followed the same analysis plan as Experiment 1 and included a preliminary analysis, an analysis of participants' performance on the three test trials (prior probability trials, posterior confirmation trials, and posterior disconfirmations trials), and an analysis of their ability to update probabilities. To better understand the development of reasoning about posterior probabilities, we conducted an exploratory analysis comparing participants' performance across age groups (4-, 5-, and 6-year-olds).

Preliminary Analysis. The preliminary analysis examined the main effects of order (Orders A–H, as described in the design section of the Methods), gender (male vs. female), age (treated as a continuous variable), and trial number (Trials 1–8). Initially, a Generalized Linear Mixed Effects Model (GLMM) was fitted to predict the probability of a correct response from the fixed effects of order, gender, age, and trial number, with participant ID included as a random effect.

There was no significant main effect of Orders A, B, and D (all $p > .10$), with only order C showing a negative effect on participants' responses ($\hat{\beta}$ = -1.04, SE = 0.49, z = -2.11, p = 0.03). Similarly, there was no main effect gender ($\hat{\beta}$ = 0.34, SE = 0.42, z = 0.81, p = 0.41), age ($\hat{\beta}$ = 0.39, SE = 0.64, z = 0.61, p = 0.54) or trial number ($\hat{\beta}$= 0.04, SE = 0.07, z = 0.49, p = 0.61). Thus order, trial number, age, and gender did not significantly influence participants' responses.

Test Trials Analysis. Chance was set at 50%, as participants could choose between two different colored blocks on a given test trial. Wilcoxon signed-rank tests were used to compare participants' performance against chance.

The average accuracy on the prior trials was 71% (SD = 0.27), significantly above chance (V = 95.5, p = 0.002, r = 0.63). The average accuracy for the posterior confirmation trials was 96% (SD = 0.14), also significantly above chance (V = 253, p < 0.001, r = 0.96). Finally, the average accuracy for the posterior disconfirmation trials was 50% (SD = 0.42), which did not differ from chance (V = 68, p = 0.51, r = 0; see Table 1).

Overall, the results indicate that participants performed above chance on the prior trials and posterior confirmation trials, but not on the posterior disconfirmation trials.

Probability Updating Analysis. In the probability updating analysis, we examined how often participants answered both the prior trials and their corresponding posterior trials correctly to assess their ability to update probabilities accurately. As in Experiment 1, chance was set at 25%.

Wilcoxon sign-ranked tests were used to compare participants' responses in the trial pairs (each consisting of a prior probability trial and its corresponding posterior probability trial) to chance (25%). Across all four sets of trials, participants answered both trial pairs correctly 47% (SD = 0.24) of the time, significantly above chance (V = 146, p < 0.001, r = 0.73).

Looking only at trial pairs that included a posterior confirmation trial, the accuracy rate was 65% (SD = 0.48), significantly above chance (V = 146, p < 0.001, r = 0.73). In trial pairs that included a posterior disconfirmation trial 29% (SD = 0.46), which did not differ from chance (V = 157, p = 0.42, r = 0.005). Thus, participants were able to successfully update probabilities, but their success was driven by their performance in the posterior confirmation trial pairs.

We then examined posterior probability trial performance in participants who scored 75% or better on the prior probability trials (i.e., correctly answered three or more out of the four prior trials). Of the 24 participants, 13 (54%) met this criterion. Wilcoxon signed-rank tests were used to compare participants' performance against chance (50%).

For these 13 participants, the average accuracy on the prior trials was 93% (SD = 0.12), significantly above chance (V = 91, p < 0.001, r = 0.91). For the posterior confirmation trials, the average accuracy was 96% (SD = 0.14), also significantly above chance (V = 78, p < 0.001, r = 0.96). For the posterior disconfirmation trials, the average performance was 34% (SD = 0.42), which did not differ significantly from chance (V = 16.5, p = 0.90, r = 0.35).

These results suggest that participants who performed well in the prior trials were able to accurately update probabilities in the posterior confirmation, but not in the posterior disconfirmation trials.

Exploratory Analysis. To assess whether children adjusted their guesses when presented with new information, we analyzed their performance across prior trials, posterior confirmation trials, and posterior disconfirmation trials in 4-, 5-, and 6-year-olds. Specifically, we examined whether children updated their responses in the two types of posterior trials.

Three separate GLMERS were fitted to predict responses in each trial type with age (treated as a continuous variable) as a fixed effect, and a random intercept for participant ID. Age did not significantly predict accuracy in the prior trials ($\hat{\beta}$= -0.30, SE = 0.48, z = -0.63, p = 0.53), or in the posterior confirmation trials ($\hat{\beta}$ = 0.34, SE = 1.52, z = 0.23, p = 0.82). However, age had a marginally significant effect in the posterior disconfirmation trials ($\hat{\beta}$ = -0.42, SE = 1.06, z = -0.39, p = 0.069), with worse performance in younger children.

To further examine whether children adjusted their guesses when probabilities changed, we compared their performance in the posterior confirmation trials and posterior disconfirmation trials to their prior trial performance. Separate GLMERs were conducted for each age group (4-, 5-, and 6-year-olds), with trial type (prior vs. posterior confirmation or prior vs. posterior disconfirmation) as a fixed effect.

When comparing the prior trials to posterior confirmation trials, children in all groups performed significantly better in the posterior confirmation trials, indicating that they adjusted their guesses in response to updated probabilities: 4-year-olds ($\hat{\beta}$= -2.39, SE = 0.79, z = -3.06, p = 0.002), 5-year-olds ($\hat{\beta}$= -1.70, SE = 0.63, z = -2.68, p = 0.007), and 6-year-olds ($\hat{\beta}$= -2.62, SE = 1.04, z = -2.50, p = 0.01).

When comparing the prior trials to the posterior disconfirmation trials, only 4-year-olds showed a significant effect of trial type ($\hat{\beta}$= 8.87, SE = 3.65, z = 2.42, p = 0.01), suggesting they struggled to adjust their guesses when probabilities changed. No significant differences were found for 5-year-olds ($\hat{\beta}$= 0.19, SE = 0.38, z = 0.52, p = 0.60), or 6-year-olds ($\hat{\beta}$= 0.29, SE = 0.41, z = 0.69, p = 0.48).

Discussion

The results of Experiment 2 suggest emerging success in probability updating. In prior probability trials, without shape information, children guessed the more probable color. In posterior confirmation trials, they updated their judgment based on new shape information, but not in posterior disconfirmation trials. Only about half consistently succeeded in prior probability trials (i.e., correctly guessing at least 3 of 4). Among those, updating occurred in posterior confirmation but not posterior disconfirmation trials. Thus, while 4-year-olds made probabilistic guesses, they struggled to revise them when faced with contradictory information.

General Discussion

Across the two experiments, we found that 4- to 6-year-old children can successfully reason about prior and posterior probabilities and integrate new information into their decision-making. This suggests that, contrary to previous findings, children as young as four can reason about posterior probabilities.

Success among 5- to 6-year-olds replicates findings from Girotto and Gonzalez (2008). Children made an accurate prior probability guess when presented only with the population and updated their responses when new information confirmed (posterior confirmation) or contradicted (posterior disconfirmation) their initial guess. Importantly, results from Experiment 2 indicate that even 4-year-olds can make probabilistic guesses and update them when provided with new evidence. Although they struggled in the posterior disconfirmation trials, their success in the posterior confirmation trials suggests they can integrate new information into their decision-making.

The results from 4-year-olds contrast with those from Girotto and Gonzalez (2008), where 4-year-olds perform at chance levels on both prior and posterior probability guesses. A key difference in our study was the 3:1 ratio of 12 blue to 4 black blocks in the prior trials, compared to the 5:3 ratio in Girotto and Gonzalez. The larger ratio may have facilitated probabilistic computation by making the difference more salient. Furthermore, the posterior confirmation trials had a probability of 1 compared to a 4:3 ratio in Girotto and Gonzalez, which may explain the improved performance in the posterior confirmation trials. While our posterior confirmation trials did not necessarily require children to make a probabilistic guess, it allowed us to assess whether children could attend to and integrate new information into their decision-making. Children did significantly better in the posterior confirmation trials than in the prior trials, suggesting that they integrated information about the block's shape.

We replicated 4-year-olds' failure in the posterior disconfirmation trials. One possibility is that 4-year-olds have difficulty switching their responses. Another possibility is that children struggle to compute the probabilities among smaller sets. Both our study and Girotto and Gonzalez had a 3:1 ratio in the posterior disconfirmation trials, but only with 4 chips/blocks. If this set size is too small for 4-year-olds to accurately compute probabilities, then increasing the set size should improve their performance. In a follow-up study, we plan to run posterior disconfirmation trials with larger sets to determine whether increasing the ratio of blocks enhances children's probability judgments.

It is possible that 5-to 6-year-olds' success could be attributed to probability matching rather than reasoning about the posterior probabilities. Under this hypothesis, children would distribute their choices in proportion to the probability of each outcome occurring. This would result in roughly 75% accuracy rate in the posterior probability trials, 100% in the posterior confirmation trials, and roughly 75% in the posterior disconfirmation trials, which aligns with our findings. However, when analyzing the performance of children who scored three or more on the prior trials, their accuracy rates in the prior probability trials were above the performance predicted by the probability matching account. This suggests that their performance was driven by reasoning about probabilities rather than probability matching.

In conclusion, our findings suggest that 4- to 6-year-old children reason about prior and posterior probabilities, and they can update their decisions by integrating new evidence. Methodological differences such as increasing the set size may explain variations from previous studies. Future studies should explore whether increasing set size improves 4-year-olds' performance in the posterior disconfirmation trials. Furthermore, future studies may investigate younger children's ability to reason about posterior probabilities to better map the developmental trajectory of this reasoning ability.

Acknowledgments

We thank the members of the Berkeley Early Learning Lab and the Computation and Language Lab for helpful discussions.

References

Denison, S., & Xu, F. (2019). Infant Statisticians: The Origins of Reasoning Under Uncertainty.

Perspectives on Psychological Science, 14(4), 499-509.

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

Denison, S., & Xu, F. (2010). Twelve- to 14-month-old infants can predict single-event probability with large set sizes. *Developmental Science, 13*(5), 798–803. https://doi.org/10.1111/j.1467-7687.2009.00943.x

Fontanari, L., Gonzalez, M., Vallortigara, G., & Girotto, V. (2014). Probabilistic cognition in two indigenous Mayan groups. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 111*(48), 17075–

Girotto, V., Fontanari, L., Gonzalez, M., Vallortigara, G., & Blaye, A. (2016). Young children do not succeed in choice tasks that imply evaluating chances. *Cognition*, *152*, 32–39. https://doi.org/10.1016/j.cognition.2016.03.010

Girotto, V., & Gonzalez, M. (2008). Children's understanding of posterior probability. *Cognition*, 106(1), 325-344. https://doi.org/10.1016/j.cognition.2007.02.005

Kushnir, T., Xu, F., & Wellman, H. M. (2010). Young children use statistical sampling to infer the preferences of other people. *Psychological Science, 21*(8), 1134–1140. https://doi.org/10.1177/0956797610376652

Ma, L. and Xu, F. (2011) Young children's use of statistical sampling evidence to infer the subjectivity of preferences. *Cognition* 120, 403–411

Pighin, S., Girotto, V., & Tentori, K. (2017). Children's quantitative Bayesian inferences from natural frequencies and number of chances. *Cognition*, *168*, 164–175. https://doi.org/10.1016/j.cognition.2017.06.028

Placì S, Fischer J, Rakoczy H. 2020 Do infants and preschoolers quantify probabilities based on proportions? R. Soc. Open Sci. 7: 191751.

Schulz, L. E., Bonawitz, E. B., & Griffiths, T. L. (2007). Can being scared causes tummy aches? Naive theories, ambiguous evidence, and preschoolers' causal inferences. Developmental Psychology, 43(5), 1124–1139. https://doi.org/10.1037/0012-1649.43.5.1124

Téglás, E., Vul, E., Girotto, V., Gonzalez, M., Tenenbaum, J. B., & Bonatti, L. L. (2011). Pure reasoning in 12-month-old infants as probabilistic inference. *Science (New York, N.Y.)*, *332*(6033), 1054–1059. https://doi.org/10.1126/science.1196404

Téglás, E., Girotto, V., Gonzalez, M., & Bonatti, L. L. (2007). Intuitions of probabilities shape expectations about the future at 12 months and beyond. *Proceedings of the National Academy of Sciences*, 104(48), 19156 19159.https://doi.org/10.1073/pnas.0700271104

Wellman, H.M., Kushnir, T., Xu, F., & Brink, K. A. (2016). Infants Use Statistical Sampling to Understand the Psychological World. Infancy, 21(5),668678.https://doi.org/10.1111/infa.12131

Xu, F. and Garcia, V. (2008) Intuitive statistics by 8-month-old infants. *Proc. Natl. Acad. Sci.* 105, 5012–5015

Yeung, H. H., Denison, S., & Johnson, S. P. (2016). Infants' Looking to Surprising Events: When Eye-Tracking Reveals More than Looking Time. *PloS one*, *11*(12), e0164277. https://doi.org/10.1371/journal.pone.0164277

Zhu, L., & Gigerenzer, G. (2006). Children can solve Bayesian problems: The role of representation in mental computation. *Cognition, 98*(3), 287–308. https://doi.org/10.1016/j.cognition.2004.12.003