Cognition 152 (2016) 32–39

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Young children do not succeed in choice tasks that imply evaluating chances

Vittorio Girotto [a,*], Laura Fontanari [a], Michel Gonzalez [b], Giorgio Vallortigara [c], Agnès Blaye [b]

[a] *Center for Experimental Research in Management and Economics, DCP, University IUAV of Venice, Italy*
[b] *Laboratory of Cognitive Psychology, CNRS and Aix-Marseille University, France*
[c] *Center for Mind/Brain Sciences, University of Trento, Italy*

ARTICLE INFO

Article history:
Received 28 October 2015
Revised 7 March 2016
Accepted 9 March 2016

Keywords:
Probabilistic cognition
Proportional reasoning
Preschoolers
Infants
U-shape development

ABSTRACT

Preverbal infants manifest probabilistic intuitions in their reactions to the outcomes of simple physical processes and in their choices. Their ability conflicts with the evidence that, before the age of about 5 years, children's verbal judgments do not reveal probability understanding. To assess these conflicting results, three studies tested 3–5-year-olds on choice tasks on which infants perform successfully. The results showed that children of all age groups made optimal choices in tasks that did not require forming probabilistic expectations. In probabilistic tasks, however, only 5-year-olds made optimal choices. Younger children performed at random and/or were guided by superficial heuristics. These results suggest caution in interpreting infants' ability to evaluate chance, and indicate that the development of this ability may not follow a linear trajectory.

1. Introduction

Are young children able to make correct probabilistic evaluations? Infants possess probabilistic intuitions, as shown by measuring their reactions to simple physical processes, such as the exit of an object from a container: the more unlikely the exit, the more the infants tend to look at it, which is typically a sign that the event is unexpected. For example, given an urn in which three identical objects and one different in shape and color bounce randomly, 12-month-olds look longer at the display when the singleton, rather than one of the identical objects, exits from the hole at the base of the urn (Teglas, Girotto, Gonzalez, & Bonatti, 2007; Teglas et al., 2011). From the age of about 5 years, children solve judgment and choice tasks in which they have to compare the chances of two competing outcomes (for a review, see Reyna & Brainerd, 1994). For example, 5-year-olds correctly predict that one is likely to get a yellow chip, if one draws a chip at random from a bag containing 3 yellow chips and 1 blue chip (e.g., Brainerd, 1981; Girotto & Gonzalez, 2008). In other words, they make correct predictions on the basis of prior possibilities. Their correct responses do not imply that they make an explicit, numerical evaluation of the chances favoring each outcome. Yet, they suggest that young children possess a basic probabilistic knowledge. From the age of about six, children solve more difficult problems in which they have to consider additional information (e.g., Girotto & Gonzalez, 2008). For example, they correctly predict that one is likely to get a yellow chip, if one draws a chip at random from a bag containing 5 yellow chips (of which 4 are square and 1 is round) and 3 blue chips (all of which are round). Then, if they are informed that one has drawn a round chip, 6-year-olds revise their evaluation, and correctly predict that one is likely to get a blue chip. Likewise, from the age of about six, children solve probability problems in which they have to consider combinations of possibilities (Gonzalez & Girotto, 2011). For example, given a bag containing various pairs of chips, each pair having a different color, 6-year-olds correctly predict that one is likely to get two chips of different colors, if one draws two chips at random from the bag. Even preliterate and prenumerate adults are able to solve problems of this sort, and their performance is similar to that of Western, educated controls (Fontanari, Gonzalez, Vallortigara, & Girotto, 2014). Taken together, these results indicate that all individuals, regardless of their instruction and culture, share the ability to infer the probability of an event extensionally, that is, by considering the different possible ways in which it may occur (Johnson-Laird, Legrenzi, Girotto, Legrenzi, & Caverni, 1999).

In between infancy and five years of age, however, little is known about young children's probabilistic competence. The available evidence is not encouraging: 3- and 4-year-olds fail not only challenging problems that require the use of posterior evidence,

---

* Corresponding author at: DCP, Convento delle Terese, 30123 Venice, Italy.
E-mail address: vgirotto@iuav.it (V. Girotto).

http://dx.doi.org/10.1016/j.cognition.2016.03.010
0010-0277/

proportions or combinatorial procedures (e.g., Girotto & Gonzalez, 2008; Gonzalez & Girotto, 2011; Piaget & Inhelder, 1975; Siegler, 1981), but also simple tasks in which they only have to consider prior information in order to predict an uncertain event. In particular, they perform poorly in simple tasks, analogous to those that have been used to investigate infants' probabilistic intuitions. For example, given a bag containing 3 yellow chips and 1 blue chip, 4-year-olds answer at chance level, if they have to predict whether a randomly drawn chip will be yellow or blue (Girotto & Gonzalez, 2008). Likewise, 3-year-olds perform at chance level, if they have to predict whether a ball, bouncing inside a rectangular box with one hole on one side and three holes on the opposite side, will exit from the one-hole side or from the three-hole one (Teglas et al., 2007/Study 3).

In sharp contrast with these negative findings, some studies have reported evidence that human infants (Denison & Xu, 2010, 2014) as well as non-human primates (Racoczy et al., 2014) possess probabilistic competence, by using choice tasks that imply evaluating chances. For example, given a jar containing 40 preferred and 10 non-preferred tokens, and a jar containing 10 preferred and 40 non-preferred tokens, 10- to 14-month-olds search in the place that hides a token drawn from the jar containing the larger number of preferred tokens (Denison & Xu, 2010, 2014). In other words, infants make optimal choices by selecting the set which is more likely to yield a preferred token. Infants make optimal choices even when they have to consider proportions. Thus, given a jar containing 16 preferred and 4 non-preferred tokens, and a jar containing 24 preferred and 96 non-preferred tokens, 10- to 13-month-olds search in the place that hides a token drawn from the jar containing the larger proportion but not the greater number of preferred tokens (Denison & Xu, 2014). In sum, preverbal participants appear to succeed in tasks in principle more demanding than those in which 3- and 4-year-olds fail. These apparently conflicting sets of results question the nature and limits of preschoolers' probabilistic cognition, and whether the development of this ability follows a linear trajectory. These questions, in turn, point to the need to test preschoolers in tasks similar to those that have provided evidence of infants' probability understanding.

Here we report three studies designed to address these questions. We used procedures inspired by the procedure designed by Denison and Xu (2010), were respondents see two sets containing various proportions of two sorts of tokens (one more attractive than the other), and must choose which set is more likely to yield an attractive token. In Denison and Xu's (2010) study, infants completed only one task. By contrast, in Study 1, children completed a series of increasingly demanding tasks. In the most elementary one, children were not required to form any probabilistic expectations because there was no uncertainty as to the outcome that each set could produce. In the most demanding tasks, in order to make optimal choices, children had to apply proportional reasoning, and resist choosing a set on the basis of superficial heuristics. Study 1 has been conducted as an independent extension and complement of Denison and Xu's (2010) study, before the publication of Denison and Xu's (2014) one, in which infants completed tasks of various difficulty levels. Study 2 investigated children's probabilistic cognition using an experimental procedure as close as possible to the one used by Denison and Xu (2014) with infants. Finally, Study 3 tested a possible alternative interpretation of the results obtained in Studies 1 and 2.

2. Study 1

In Study 1, 3- to 5-year-olds completed the tasks depicted in Fig. 1. In the simplest one (Task A), both outcomes were certain. Hence, in order to make an optimal choice, children had simply to distinguish the two sets, with no need to form probabilistic expectations. In the intermediate tasks (Tasks B), only the favorable set could yield a certain outcome. In one version (Task B1), the favorable set contained a greater number of attractive tokens than the unfavorable set. Hence, children could make optimal choices either by attributing the certainty of yielding an attractive token to the favorable set or, more simply, by using the *absolute number heuristic*, that is, selecting the set containing the greater number of attractive tokens. In another version (Task B2), the favorable set contained a smaller number of attractive tokens than the unfavorable one. Unlike the former version, in this version children had to resist selecting a set on the basis of the absolute number heuristic to make an optimal choice. In the most demanding tasks (Tasks C), both outcomes were uncertain. In one version (Task C1), the favorable set contained a greater number of attractive tokens than the unfavorable one. Hence, children could make optimal choices either by distinguishing the two sets according to their respective ratios of attractive and unattractive tokens hence, forming probabilistic expectations as to which set was more likely to yield a positive outcome or, more simply, by applying the absolute number heuristic. In another version (Task C2), the favorable set contained a smaller number of attractive tokens than the unfavorable one. Hence, children had to resist selecting a set on the basis of the absolute number heuristic to make an optimal choice.

2.1. Method

2.1.1. Participants

In Studies 1 and 2, participants were children attending public preschools in Trento (Italy). Their participation was approved by a signed consent obtained from parents. In Study 1, we tested 93 children (41 girls) distributed into three age groups: 3-year-olds ($n = 35$; mean age: 3.52; range: 2.98–4.00), 4-year-olds ($n = 33$; mean age: 4.67; range: 4.01–5.00), and 5-year-olds ($n = 25$; mean age: 5.60; range: 5.01–6.01). We tested 2 further children, but we did not consider their answer because they failed to understand the instructions.

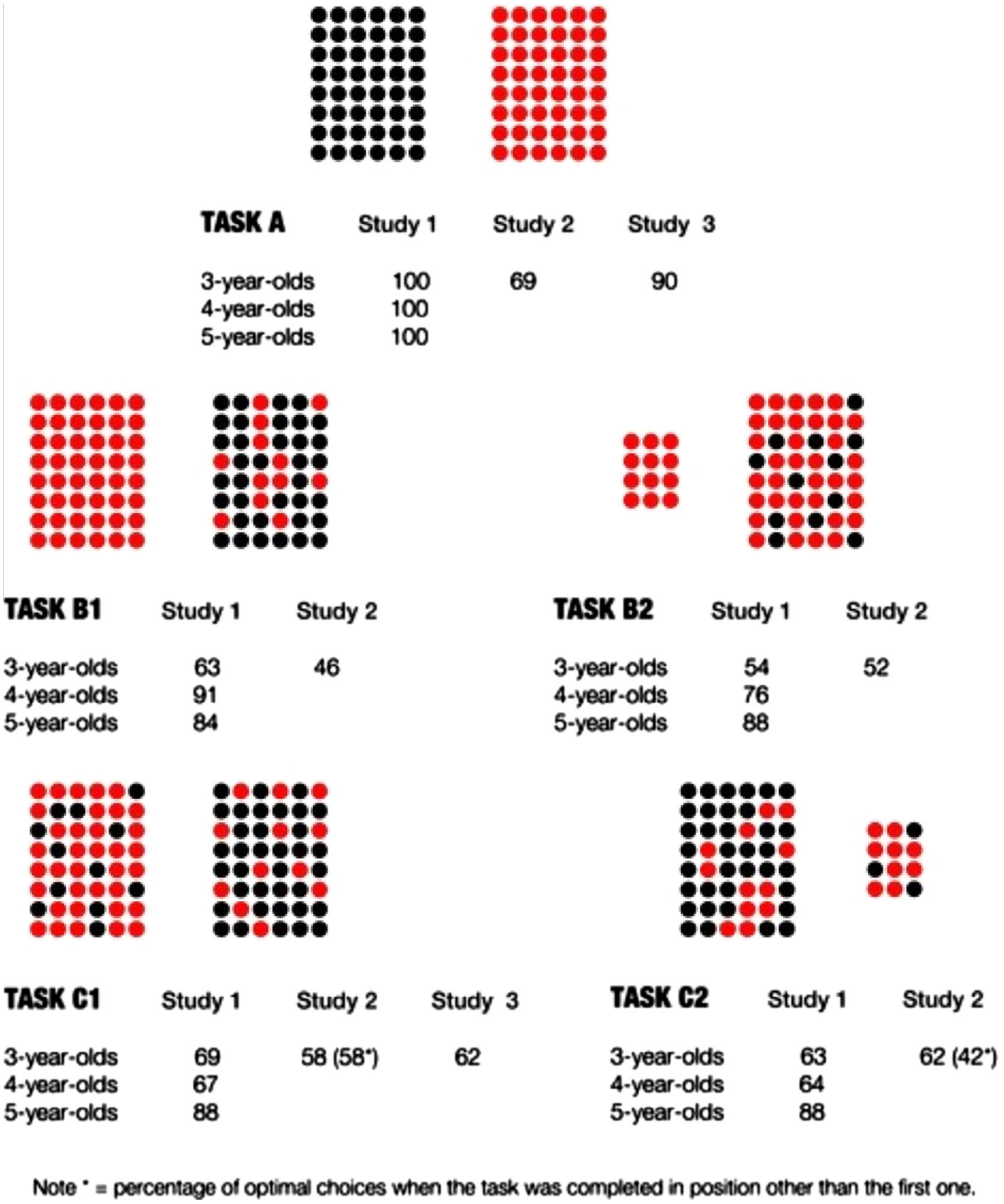
2.1.2. Materials, procedure and design

Each child was tested individually, in a quiet room. The experimenter informed the children they would play games in which they could win some stickers. Children sat in front of a table upon which two opaque boxes (30 × 15 × 10 cm) were placed. Each box and a circular hole (10 cm in diameter) on its top, two groups of wooden chips (2 cm in diameter) colored in red or black, and two cardboards (29 × 21 cm), each depicting one group of chips. The boxes were placed approximately 30 cm apart. In front of each box, there was an opaque mug (10 cm in height, 4 cm in diameter). To start, the experimenter explained the rules of the game:

"We will play with these two puppets [the experimenter named and pointed at two animal-toys: an elephant and a koala]. Which one do you prefer? [The child chose one puppet] OK. Now, you and your puppet will belong to the red team. This red sticker is for you, and this red sticker is for your puppet [the experimenter distributed the stickers]. The other puppet belongs to the black team. So, I will give him this black sticker [the experimenter placed a black sticker on the other puppet].

Now, we will play with some red and some black chips [the experimenter showed some chips]. The red chips make your red team win the game. The black chips make the black team win the game. Every time you find a red chip, your red team wins a sticker. Every time you find a back chip, the black team wins a sticker."

The experimenter checked whether the children understood the instructions by asking them to name the winning color and to point to a winning chip. If they failed, she corrected them. Then, she went on:



TASK A	Study 1	Study 2	Study 3
3-year-olds	100	69	90
4-year-olds	100		
5-year-olds	100		

TASK B1	Study 1	Study 2
3-year-olds	63	46
4-year-olds	91	
5-year-olds	84	

TASK B2	Study 1	Study 2
3-year-olds	54	52
4-year-olds	76	
5-year-olds	88	

TASK C1	Study 1	Study 2	Study 3
3-year-olds	69	58 (58*)	62
4-year-olds	67		
5-year-olds	88		

TASK C2	Study 1	Study 2
3-year-olds	63	62 (42*)
4-year-olds	64	
5-year-olds	88	

Note * = percentage of optimal choices when the task was completed in position other than the first one.

Fig. 1. Percentage of optimal choices (i.e., selecting the set which is more likely to yield a red, winning token) by task and age group in the three studies. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

"See these chips? [The experimenter pointed to one group of chips.] We will put them into this box. [The experimenter pointed to one box, and encouraged the child to help her in filling the box with the mentioned chips] To help you remember the chips that we put into the box, I made a drawing of these chips, here. [The experimenter pointed to one cardboard, representing the content of the first box, and placed it in front of the box.] Now, see these other chips? [The experimenter pointed to the other group of chips] We will put them into this box. [The experimenter pointed to the other box, and filled it with the child's help] To help you remember the chips that we put into the box, I made a drawing of these chips, here. [The experimenter pointed to the other cardboard, representing the content of the second box] Now, without looking, I will take one chip from this box. [The experimenter pointed to one box] And I will take one chip from this box. [The experimenter pointed to the other box.] And, without looking, I will put these chips in these mugs. [The experimenter took one chip from each box and placed it in the corresponding mug.] Now, you choose one mug. Remember that, to win, you have to find a red chip. If you find a black chip, the black team wins. Before making your choice, you may look at these two drawings. [The experimenter pointed to the two cardboards.] So, which mug do you choose to find a red chip?"

The children made their choice, by naming or pointing to one mug. The experimenter opened it and assigned the prize. She then moved to the next task, saying: "Now we play with these new boxes and these new chips." The children completed the tasks in the order of complexity indicated in Fig. 1. In all of Denison and Xu's (2010, 2014) studies, half of the infants complete a 'no switch' task in which the experimenter placed the drawn objects into the mugs adjacent to the jars from which they were removed. The other half of the infants completed a 'switch' task in which the experimenter placed the objects drawn into the mugs opposite to the jars from which they were removed. Infants performed above chance level with both sorts of task. For this reason, in Study 1, we used only 'no switch' tasks, because this procedure was easier to implement. The order in which the experimenter introduced and filled the two boxes (favorable vs. unfavorable) and the position (left vs. right) of the favorable box were counterbalanced.

2.2. Results

In Task A, were both outcomes were certain, all children selected the mug that contained a reward, binomial tests, $p < .0001$ (see Fig. 1). In the other tasks, however, the youngest children performed poorly. In Task B1, 3-year-olds did not perform better than chance: only 22 out of 35 selected the mug that certainly contained a reward, $p < .09$; 95% Confidence interval [46,77]. By contrast, both 4-year-olds (correct: 30 out of 33) and 5-year-olds (correct: 21 out of 25) performed better than chance: $p < .0001$; 95% Confidence interval [76,98], and $p < .0005$; 95% Confidence interval [65,94], respectively. Likewise, in Task B2, the youngest children (correct: 19 out of 35) did not perform above chance level, $p > .37$; 95% Confidence interval [38,70]. By contrast, 4-year-olds (correct: 25 out of 33) and 5-year-olds (correct: 22 out of 25) selected the favorable mug reliably different from chance: $p < .003$; 95% Confidence interval [59,87], and $p < .0001$; 95% Confidence interval [69,97], respectively. In one of the most demanding tasks (Task C1), all children's groups performed reliably better than chance: 3-year-olds (correct: 24 out of 35), $p < .03$; 95% Confidence interval [52,81]; 4-year-olds (correct: 22 out of 33): $p < .05$; 95% Confidence interval [50,80]; 5-year-olds (correct: 22 out of 25): $p < .0001$; 95% Confidence interval [69,97]. However, in the task in which they had to resist choosing a mug on the basis of the absolute number heuristic (Task C2), both 3-year-olds (correct: 22 out of 35) and 4-year-olds (correct: 21 out of 33) did not perform reliably better than chance: $p > .08$; 95% Confidence interval [46,77], and $p > .08$; 95% Confidence interval [47,78], respectively. By contrast, 5-year-olds' (correct: 22 out of 25) performance reliably differed from chance: $p < .0001$; 95% Confidence interval [69,97]. Youngest children performed above chance level on Task C1. Their performance on this task, however, was not significantly better than their performance on Task B1 (8 children fit this pattern, 5 went against it, and the rest were ties; sign test, $p > .38$), or on Task B2 (12 children fit this pattern, 7 went against it, and the rest were ties; sign test, $p > .18$). It should be noted that this is the only case in which children performed above chance level on a demanding task, and did not perform better than chance on less demanding ones.

Together, these results fail to provide evidence that 3- and 4-year-olds use probability and proportion evaluations in dealing with simple choice tasks. Their failure cannot be attributed to a misunderstanding of the logic of the task. Indeed, on the certainty task (Task A), they performed correctly. By contrast, on the probability tasks, they perform randomly or guided by the absolute quantity heuristic. What might account for the difference between their poor performance and the good performance exhibited by infants in the same sort of tasks? One possibility is that the procedure we used to investigate children's probabilistic ability differed in some critical aspects from the one that have elicited good performance in infants. We addressed this possibility in Study 2.

3. Study 2

Study 2 was designed to replicate Denison and Xu's (2010, 2014) procedure. Two aspects that characterized the approach we used in Study 1 might have hindered young children's performance. First, Denison and Xu's participants had to move toward the preferred mug. Our participants had to indicate or to point to the preferred mug. Second, Denison and Xu's participants made a choice in order to get a reward. Our participants made a choice in order to get a red chip, which implied getting the reward. One possibility is that young children have difficulty to express their expectancies by means of a verbal answer or a pointing gesture, and/or to understand the link between an arbitrary token and a reward. Accordingly, in Study 2 we asked children to move toward a mug, rather than naming or indicating it, and to search directly for a reward, rather than for a token representing a reward.

Notice that the procedure employed in Study 1 differed from Denison and Xu's one in three further points, which, in principle, could facilitate young children's performance. First, Denison and Xu's infants were presented with two jars that were already filled, and their contents were not entirely visible. By contrast, in Study 1, children were presented with two empty boxes, and witnessed the experimenter placing the chips in each box. Second, in the choice phase of Denison and Xu's studies, infants could not see the content of the jars and had to rely on their memory to recall what the jars contained. By contrast, in the choice phase of Study 1, children saw the cardboards that represented all the chips contained in each box. Third, Denison and Xu's infants did not complete any familiarization task before tackling the probability one. By contrast, in Study 1, before tackling the probability tasks, children completed the certainty task, in which all participants made the suitable choice and received the reward. Despite the potential benefits of these procedural features in Study 1, these were not used in Study 2. Therefore, children only saw some of the objects contained in each set, made their choice without the mnemonic aid provided by the pictorial representation of the two sets, and did not complete any task before the probability ones.

In Study 1, the older children performed well in all tasks, including those in which they had to consider proportions. For this reason, in Study 2 we tested only 3- and 4-year-olds. They had to complete a series of tasks, just as in Study 1. Denison and Xu's infants, however, completed only one task. Thus, to facilitate comparison between studies, Study 2 participants started with one of the two tasks involving proportions (Task C1 or else Task C2). Then, they completed the other proportional reasoning task as well as the remaining ones (A, B1 and B2). If 3- and 4-year-olds are able to apply correct probabilistic evaluations in this sort of tasks, they should succeed in the most challenging ones, just as 10–14-month-olds appear to do.

3.1. Method

3.1.1. Materials

We could not use lollipops as rewards because the school where we conducted the study forbade providing food to children. Thus, we used stickers as rewards, and plastic spoons covered with colored paper as stimuli. Some spoons were rewarding because they had a sticker on their bowl. Some other ones were not rewarding because they had nothing on their bowl (see Supplementary Information, Pictures). We used spoons as stimuli because the spoons' hands, just as the lollipops' sticks, allowed children to see that only one token was removed from each jar during the drawing phase of the test (see below). On the two proportional reasoning tasks (C1 and C2), the colors of the spoons were similar to those of the lollipops used by Denison and Xu (i.e., pink and black), namely: fuchsia and black or else carmine and blue. The rewarding spoons had always the brighter color (i.e., fuchsia, carmine). In order to maintain children's interest in the game, we used a couple of different colors in each task (see Supplementary Information, Pictures).

3.1.2. Participants and procedure

The participants were 48 children (20 girls; mean age = 3.79; range = 3.07–4.4). They sat on the floor at about 2 m in front of the experimenter. Each child was tested individually in the five tasks detailed in Fig. 1.

3.1.2.1. Preference task. As in Denison and Xu's (2010, 2014) studies, prior to each test task, the participants completed a preference task. The experiment presented two spoons, one of each sort, and drew the child's attention to each one: "See this spoon? It has a sticker on it. See this one? It has nothing on it." Next, the experimenter drew the child's attention to both spoons at the same time, placed them on the floor, at equal distance from the child, and said: "Do you want to come and pick up the spoon you prefer?" The experimenter clapped for his/her choice, commenting: "Good job! You found the spoon you prefer!"

3.1.2.2. Test tasks. Once the experimenter had established children's preference regarding the spoons, she presented two transparent but covered jars (17 cm in height, 10 cm in diameter), and two opaque mugs (10 cm in height, 8 cm in diameter). Each mug had an opaque cover. The experimenter simultaneously removed the cover from the two jars and showed their respective contents. On tasks B1, B2 and C2, the content of the jars varied as a function of children's preferences. For example, suppose that on the preference test of Task C2 a child chose the spoon without the sticker. In this case, the experimenter presented a couple of jars in which one jar contained a greater proportion of spoons without stickers. By contrast, on tasks A and C1, all children received the same two jars because, regardless of what their preference was, the jar containing a greater proportion of the spoons that they preferred was the favorable jar (see Fig. 1). Next, the experimenter lifted one jar off the floor, shook and rotated it. Then, she did the same with the other one. In the drawing phase, the experimenter lowered the cover over the two jars simultaneously, closed her eyes, drew one spoon from one jar and, covering its bowl in her hand, put it into the mug next to the jar. Finally, she closed the cover of the mug, without letting the child seeing the color of drawn spoon. She did the same with the other jar. Next, in the choice phase, she lifted both mugs simultaneously, inviting the child to move toward the mugs: "Come choose one." To make the task clearer, the experimenter added: "You can keep the spoon you get." In Denison and Xu's studies, the experimenter did not add this sentence. We deemed that such a clarification was unlikely to affect performance since both mugs contained a spoon. Moreover, the experiment said this to all children. Notice, by contrast, that in Denison and Xu's studies, the experimenter modified the instructions as a function of the initial reaction of the infant: "For the more hesitant infants, the experimenter sometimes had to show the lollipops twice on the preference trial, or move them closer to the infant" (Denison & Xu, 2014, p. 339). Once the child made his/her choice by moving toward the mugs (see Supplementary Materials, Pictures), the experimenter opened the chosen mug and gave the spoon contained in it to the child. Then, she presented the other tasks.

In Denison and Xu (2010), the experimenter shook and rotated the jar on the left (on the right in Denison & Xu, 2014), and drew one object from it. Then, she shook and rotated the jar on the right (on the left in Denison & Xu, 2014), and drew one object from it. In Study 2, the order in which the experimenter lifted, shook and rotated the jars, and drew the spoons was counterbalanced. In Study 1, children completed only the 'no switch' version of each task (i.e., the experimenter placed each drawn token into the mug that was next to the box from which it had been removed). Nonetheless, 3- and 4-year-olds failed. For this reason, in Study 2, we continued to use only the 'no switch' procedure. Half of the children started with Task C1, in which the favored jar contained a greater amount as well a greater proportion of rewards. The other half started with Task C2, in which the favored jar contained a greater amount but not a greater proportion of rewards. The remaining four tasks were presented in a random order. The order of presentation of the two sorts of spoon (rewarding vs. non rewarding), and the position (left vs. right) of the favorable jar was counterbalanced.

3.2. Results and discussion

In the crucial tasks presented initially, children did not display a reliable preference for the favorable mug (see Fig. 1). In C1 (correct: 14 out of 24) and C2 (correct: 15 out of 24) tasks, they did not perform better than chance, $p > .26$; 95% Confidence interval [39,76], and $p > .14$; 95% Confidence interval [43,79], respectively. A similar pattern emerged when children completed the crucial trials in position other than the first one. In C1 task, their performance (correct: 14 out of 24) did not differ from chance, $p > .26$; 95% Confidence interval [39,76]. On C2 task, their performance (correct: 10 out of 24) was inferior to the chance level. In the remaining tasks too, children's performance did not suggest the presence of probabilistic evaluations. In Task A, they selected the favorable mug reliably different from chance (correct: 33 out of 48), $p < .007$; 95% Confidence interval [55,80], suggesting that they understood the logic of the task. In the other two tasks, however, children performed poorly. In both Task B1 (correct: 22 out of 48) and B2 (correct: 25 out of 48) they did not perform better than chance, $p > .33$; 95% Confidence interval [33,60], and $p > .44$; 95% Confidence interval [38,66], respectively.

It should be noted that in Task A children's performance was better than chance but, unlike in Study 1, not at ceiling (69% correct vs. 100% correct, respectively). One possibility, pointed out by an anonymous reviewer, is that children did not perform at ceiling level on Task A of Study 2 because they did not have a strong preference for a given outcome (i.e., getting the brighter colored, sticker-carrying spoon vs. getting the other, empty one). In fact, most (75%) children preferred the sticker-carrying spoon. Moreover, the rate of success among children who preferred the sticker-carrying spoon did not differ significantly from that of children who preferred the empty spoon (67% vs. 75% correct, respectively, $\chi^2 = 0.29$, $p > .59$). Another possibility is that children did not perform at ceiling level on Task A of Study 2 because of the order in which they completed the various tasks. Unlike in Study 1, in which participants completed Task A as the initial task, in Study 2 participants completed it as the second, third, fourth or fifth task. It is noteworthy that their performance on the task decreased as a function of the task rank (85%, 69%, 64%, and 54% correct, respectively). Along with the perfect performance obtained in Study 1, this trend suggests that completing tasks whose outcome is uncertain might negatively affect young children's performance on an ensuing task whose outcome is certain. This possibility deserves future studies. In any case, in Study 3 we used a procedure such that children could not be indifferent between the two outcomes.

In sum, in a study whose procedure replicated the one used by Denison and Xu's (2014), 3- and 4-year-olds made optimal choices only on the task in which they had to compare two certain outcomes. In all other choice tasks, including the simplest ones, their performance did not provide evidence for an ability to carry out a correct evaluation of probabilities.

4. Study 3

It might be argued that 3- and 4-year olds failed the choice tasks of Studies 1 and 2 not because they lacked the ability to evaluate chance, but because they failed to conceive the random nature of the outcomes about which they had to make a choice. In Studies

1 and 2, these outcomes were the result of the experimenter's action. If 3- and 4-year-olds believed that the experimenter intended to deceive them, and might even draw some specific objects from the jars, they could not use prior possibilities to make an informed choice, and so they performed at random. Indeed, sensitivity to sampling conditions emerges early in the course of development (Xu & Denison, 2009). Notice that in Denison and Xu's (2010, 2014) studies, infants were in the same condition as our young children, namely, they observed the experimenter performing the drawing process and made a choice about its outcomes. Yet, unlike young children, infants succeeded. This piece of evidence, however, does not rule out the proposed interpretation because, unlike young children, infants might fail to perceive deceiving intentions in others' actions (see e.g., Wellman & Liu, 2004). Therefore, in order to provide evidence that young children did not misconceive the nature of the drawing process, we conducted a study in which 3-year-old children chose between two outcomes that had been produced by a mechanical device, rather than by a human agent's action.

Infants appear to reason correctly when the ratios of winning to non-winning tokens in the competing populations are: 3:1 versus 1:3 (Denison & Xu, 2014), as well as when the ratios are: 4:1 versus 1:4 (Denison & Xu, 2010). In Studies 1 and 2, children completed tasks involving the ratios 3:1 versus 1:3. Thus, in order to generalize our results, in Study 3 we employed the ratios 4:1 versus 1:4. All children completed the certainty task (Task A) and then one of the tasks involving proportions (Task C1).

4.1. Method

4.1.1. Participants, materials and procedure

The participants were 21 children (10 girls; mean age = 3.75; range = 2.92–4.08), attending public preschools in Aix-en-Provence (France). Each child was tested individually in two tasks. The experimenter informed the children they would play games and could win some balls and stickers. They sat in a chair at about 1.5 m in front of the experimenter, who sat on the floor. Between them there was a transparent plastic urn (of approximatively 4500 cm^3 in volume) which was open at its top and ended in an opaque neck that was inserted into a wooden box (see Supplementary Information, Study 3, Pictures). On one side of the box, there was a handle. Pulling it back and forth made one of the objects contained in the urn fall down. Under the box, there was an opaque mug (10 cm in height, 6.5 cm in diameter). Before starting the tasks, the experimenter presented the kind of stimuli (ping pong balls covered with stickers of different colors) and exemplified how six balls placed into the urn can fall down one by one into a mug each time a handle placed on the side of the device was pulled.

Next, the experimenter placed two new balls on the floor, at about 30 cm from each other. One ball was painted in black. The other one was white and covered with golden and orange stickers. The experimenter checked whether the child was able to distinguish the balls "Which is the black ball? Which is the golden ball?" If they failed, she corrected them. Then, she emphasized the beauty of the golden ball and introduced the game "How nice is the golden ball? It's really nice. Let's make a game so that you may win a golden ball."

The first trial was a version of Task A (Fig. 1) with 10 tokens in each set. The experimenter presented the children with two transparent urns. She filled one urn with 10 black balls, and the other with 10 golden balls. In each case, she commented "Look at the balls I put in this machine." She then mixed up the balls with a little stick, saying "I'm mixing up the balls, look how I do it," so that children could see the content of the urn, and she explained the game's rule:

"Look, one ball will fall from this machine down into this mug, and one ball from this machine will fall down into this mug, when I pull the handles. [The experimenter indicated one urn and the corresponding mug, then the other one and the corresponding mug] You choose a mug. If there is a golden ball in it, you keep the ball. If there is a black ball in it, I keep the ball. [Each time the experimenter mentioned a ball, she showed an exemplar of the ball she was talking about] Now I make one ball fall down from this machine. [The experimented put her hand on the top of the mentioned urn, in order to draw children's attention on the urn, rather than on the mug]. Now the ball is in this mug." While pointing to the mug, the experimenter covered it with a square cover of 12 cm in side length. Then she put the mug in front of the urn. She did the same with the other urn and mug. After the drawing phase, the experimenter recalled the game's rule to the child and invited him/her to choose one mug. The child made his/her choice, by pointing to one mug. The experimenter opened it and assigned the prize. The second trial was a version of Task C1 (Fig. 1) with 50 tokens in each set. The favorable urn was filled with 40 golden balls and 10 black balls (4:1). The other one was filled with 10 golden balls and 40 black balls (1:4). As in Studies 1 and 2, children completed only the 'no switch' version of each trial. The order in which the experimenter operated on the two urns (favorable vs. unfavorable) and the position (left vs. right) of the favorable urn were counterbalanced.

4.2. Results and discussion

In Task A, children's selection of the favorable mug was reliably different from chance (correct: 19 out of 21), $p < .0001$; 95% Confidence interval [70,99], suggesting that they understood the logic of the task. In Task C1, however, they did not perform better than chance (correct: 13 out of 21), $p > .19$; 95% Confidence interval [38,82].

In Study 2, children had to express a preference for one of two possible outcomes (i.e., getting the brighter colored, sticker-carrying spoon vs. getting the other, empty one). By contrast, in Study 3, children were not required to express any preference because one outcome was clearly positive (i.e., getting a ball) and the other one was clearly negative (i.e., getting nothing). Children did prefer the positive outcome given that they performed almost at ceiling on the certainty task. Thus, their failure to choose the favorable mug in Task C1 has to be attributed to a lack of probability understanding rather than to a lack of motivation to obtain one type of outcome.

In sum, 3-year-olds appear to be unable to reliably predict the outcome of a random sampling, even when it has been produced by a mechanical device, and not by the action of a human agent.

5. General discussion

In three studies, we found that 3–4-year-olds fail simple tasks that ask to predict the occurrence of random outcomes. Children from 3 to 5 years of age were presented with two sets of attractive and unattractive tokens. One token was randomly drawn from each set and placed into a separate opaque mug. The children's task was to choose a mug in order to get an attractive token. If children were able to make correct probability estimations, they would choose the mug with the token drawn from the set which was more likely to yield a favorable outcome. Children understood the logic of the task: In the certainty task, in which one set contained only attractive tokens and the other one contained only unattractive tokens, children of all age groups chose the mug with the token drawn from the favorable set. However, in the remaining tasks, which require forming probabilistic expectations, only 5-year-olds made optimal choices. Younger children performed randomly and/or guided by superficial heuristics, like preferring the set containing a larger number of attractive tokens rather than the set containing a greater proportion of attractive tokens. Young children's failures occurred in a variety of conditions. Children failed when they reasoned about events produced by an intentional agent (Studies 1 and 2), and about events produced by a chance set-up device (Study 3). They failed when they had to rely on their memory to recall the content of the sets (Studies 2 and 3), as well as when a pictorial representation of the sets provided them with a mnemonic aid (Study 1). They failed when the ratio of attractive to unattractive tokens in the two sets was 3:1 versus 1:3 (Studies 1 and 2), and when it was 4:1 versus 1:4 (Study 3). Finally, they failed regardless of whether they completed the probabilistic tasks after (Studies 1 and 3) or before (Study 2) tasks that did not involve probabilistic expectations.

The present results confirm and extend those obtained in previous studies on preschoolers' probabilistic cognition. Before the age of about 5 years, children err in complex tasks in which they have to consider posterior possibilities (e.g., Girotto & Gonzalez, 2008), proportions (e.g., Siegler, 1981) or combinations of possibilities (Gonzalez & Girotto, 2011). Young children, however, also err in simple tasks, analogous to those used in the present studies, in which they have to consider only prior possibilities (e.g., Girotto & Gonzalez, 2008; Teglas et al., 2007/Study 3). Along with the present results, these findings lead to negative conclusions about young children's probabilistic knowledge. Indeed, 3–4-year-olds answer randomly when they have to predict which of two outcomes is more likely to occur (Girotto & Gonzalez, 2008), as well as when they have to express their prediction by means of an overt behavior such as moving toward the place that is more likely to hide a reward (Studies 1–3). Moreover, they answer randomly when they reason about large sets of elements (Studies 1–3), but also when they reason about very few elements (Teglas et al., 2007/Study 3).

This negative picture contrasts with the one emerging from the investigation of probabilistic cognition in infants (Denison & Xu, 2010, 2014) and non-human primates (Racoczy et al., 2014). These respondents appear to be able to solve the tasks that 3- and 4-year-olds fail, including the most demanding ones, that is, those involving proportional reasoning. What may be the source(s) of this conflicting evidence? One possibly is that infants' probabilistic abilities have been overestimated. Denison and Xu (2014, p. 338) reported asking parents "to hold their infant in front of them, and to refrain from talking, pointing or influencing their child in any way." Were these instructions followed by all parents or did some of them inadvertently oriented their child's body toward the favorable set? In fact, Denison and Xu (2014, p. 337) reported that one participant was excluded because of "parental interference," which suggests that they did monitor this type of behavior. In any case, these are empirical questions that warrant future research. Only studies aimed to address them could establish whether infants' choices actually reveal their probabilistic intuitions or their parents' biased behavior.

Another possibility is that what we interpreted as overall poor performance in 3-year-olds indeed concealed two subgroups, one that would have near perfect understanding and another that would lack any probabilistic intuitions. This interpretation, however, is not supported by the reported evidence: In Study 1, only 17% of 3-year-olds solved all tasks correctly and only 6% in this age group failed on most tasks (0 or 1 success). The remaining 77% passed some tasks and failed on others (i.e., 2, 3 or 4 successes). Likewise, in Study 2, only 8% of children succeeded systematically and only 13% failed on most tasks (0 or 1 success), leaving 79% at an intermediate level of performance.

A further, more important possibility is that infants do succeed in choice tasks that imply evaluating chances, and the difference between infants and young children's choices reflects the limits of young children's inhibitory control. The procedure used in the present studies required children to accept a delay before being rewarded. They were explicitly told that if their choice turned out to be successful, they would obtain a reward. However, they were also told to sit still, and watch the experimenter drawing and putting the two tokens into the two mugs. In other words, children had to exert self-regulation in order to prevent a spontaneous tendency to go and get the reward from the containers. Inhibiting a prepotent behavior is costly for children, and develops dramatically during the preschool years (Carlson, 2005; for reviews, see Diamond, 2013; Garon, Bryson, & Smith, 2008). Accordingly, young children's random performance in the reported studies might be the result of an overload on working memory produced by tasks' inhibition requirements. Such requirements did not directly apply to infants' testing where parents held their child in front of them, and exerted the required control by letting them go only when relevant to do so. This interpretation is similar to the one proposed to explain the existence of other paradoxical tendencies in the development of physical reasoning (e.g., Baker, Gjersoe, Sibielska-Woch, Leslie, & Hood, 2011). More importantly, this interpretation can account for two sets of previously reported results. First, young children answered randomly in tasks in which they had to make predictions or choices in order to obtain a reward (Girotto & Gonzalez, 2008). Second, they performed successfully in tasks in which they had to produce a behavior without possible reward. Indeed, in a situation where children's task was to press a button when they saw a ball exiting a box, 3-year-olds acted more quickly when the ball exited from the three-hole side than from the opposite one-hole side (Teglas et al., 2007/Study 3). This finding suggests that children had prepared their response by directing their attention toward the three-side hole, that is, by anticipating the more likely outcome. And they were likely to do so because they did not need to suppress any tendency to seek a reward.

There are prediction tasks, however, which do not require any control over reward-seeking behaviors, in which young children perform poorly. Indeed, the same 3-year-olds whose motor behavior indicated the presence of probabilistic intuition answered randomly to the question of whether the ball would exit the box from the three-hole side or from the one-hole side (Teglas et al., 2007/Study 3). The inhibitory control interpretation cannot easily explain their failures. Likewise, it cannot easily explain the finding that non-human primates perform successfully in the same choice tasks in which young children err (Racoczy et al., 2014). It should be noted, however, that the apes that exhibited probabilistic abilities were adult animals (ranging in age from 6 to 30 years). Thus, if one assumes that adult apes have inhibitory abilities similar to those of the older children and greater than those of the younger children tested in the present studies (see e.g., Vlamings, Hare, & Call, 2010), one might attribute apes' success and young human children's failure to differences in inhibitory abilities. A more prudent conclusion is that research documenting the development of probabilistic cognition in apes is needed to complement the results obtained with adult animals.

Differences between infants' and young children's performance have been observed in domains other than probabilistic knowledge, like naive physics (e.g., Baker et al., 2011; Keen, 2003) and naive psychology (e.g., Baillargeon, Scott, & He, 2010). In all these cases, infants' abilities seem to disappear in early childhood. A prevalent explanation of these tendencies is that they depend on the less demanding nature of the tasks used to investigate infants' abilities as opposed to children's ones: Infant studies typically measure infants' spontaneous reaction to an unfolding scene. By contrast, child studies typically measure children's explicit response to a direct question. This explanation can plausibly account for the differences observed in other domains, but it cannot easily be applied to the probabilistic domain. Indeed, in our studies young children completed the same sort of explicit-response tasks that infants appear to succeed. Unlike infants, they failed.

Young children's failures suggest caution in interpreting infants' optimal choices in probabilistic tasks. However, these failures do not call into question the hypothesis of an early probabilistic competence because they concern only the choice paradigm. Infants' probabilistic intuitions have been documented by measuring their reactions to the outcome of physical processes (e.g., Teglas et al., 2007/Studies 1 and 2; Teglas et al., 2011). Unlike the ability to make optimal choices, these intuitions do not disappear in early childhood. Indeed, in tasks asking for a motor reaction, young children correctly anticipate the more probable outcome of a physical process, like the exit of a ball from a box (Teglas et al., 2007/Study 3).

To conclude, preverbal or barely verbal infants manifest probabilistic intuitions in their reactions to the outcomes of simple physical processes (Teglas et al., 2007/Studies 1 and 2; Teglas et al., 2011). They also seem to possess the capability to use these intuitions in their choices (Denison & Xu, 2010, 2014). Young children continue to manifest sensitivity to probabilities in their motor behavior (Teglas et al., 2007/Study 3). However, both their verbal judgments (Girotto & Gonzalez, 2008; Teglas et al., 2007/Study 3), and their explicit choices (Studies 1–3) fail to reveal the ability to evaluate chances. One possibility is that infants' probabilistic abilities have been overestimated. Another possibility is that there is a genuine U-curve in the development of probabilistic cognition: infants do make choices on the basis of probabilistic considerations. This ability disappears in the course of development because of inhibitory control deficits or because of other unknown factors, and re-emerges only at the age of about 5–6 years. Whereas the reported results cannot disentangle these two possibilities, they strongly indicate the necessity to examine in more detail the emergence of this important but hitherto neglected ability.

Acknowledgments

We thank the schools, teachers, parents and children for making this research possible, Davide Potrich, Beatrice Siliprandi and Pascal Surian for their help in collecting the data, and Jasmine Loveland Lopez for her comments. The research was supported by grants from GAM-Fondazione Ca' Foscari and the Italian Ministry of Research (PRIN2010-RP5RNM) to Vittorio Girotto, and by an ERC Advanced Grant (PREMESOR ERC-2011-ADG_20110406) to Giorgio Vallortigara. The funding agreements ensured the authors' independence in designing the study, collecting and interpreting the data, writing, and publishing the report.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.cognition.2016.03.010.

References

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

Baker, S. T., Gjersoe, N. L., Sibielska-Woch, K., Leslie, A. M., & Hood, B. M. (2011). Inhibitory control interacts with core knowledge in toddlers' manual search for an occluded object. *Developmental Science, 14*, 270–279.

Brainerd, C. J. (1981). Working memory and the developmental analysis of probability judgment. *Psychological Review, 88*, 463–502.

Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology, 28*, 595–610.

Denison, S., & Xu, F. (2010). Twelve- to 14-month-old infants can predict single-event probability with large set sizes. *Developmental Science, 13*, 798–803.

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

Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64*, 135–168.

Fontanari, L., Gonzalez, M., Vallortigara, G., & Girotto, V. (2014). Probabilistic cognition in two indigenous Mayan groups. *Proceedings of the National Academy of Sciences of the United States of America, 111*, 17075–17080.

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin, 134*, 31–60.

Girotto, V., & Gonzalez, M. (2008). Children's understanding of posterior probability. *Cognition, 106*, 325–344.

Gonzalez, M., & Girotto, V. (2011). Combinatorics and probability: Six- to ten-year-olds reliably predict whether a relation will occur. *Cognition, 120*, 372–379.

Johnson-Laird, P. N., Legrenzi, P., Girotto, V., Legrenzi, M. S., & Caverni, J. P. (1999). Naive probability: A mental model theory of extensional reasoning. *Psychological review, 106*, 62–88.

Keen, R. (2003). Representation of objects and events why do infants look so smart and toddlers look so dumb? *Current Directions in Psychological Science, 12*, 79–83.

Piaget, J., & Inhelder, B. (1975). *The origin of the idea of chance in children.*London: Routledge & Kegan Paul (Translation of original work (1951)).

Racoczy, H., Cluver, A., Sauke, L., Stofferegen, N., Grabener, A., Migura, J., & Call, J. (2014). Apes are intuitive statisticians. *Cognition, 131*, 60–68.

Reyna, V. F., & Brainerd, C. J. (1994). The origins of probability judgment: A review of data and theories. In G. Wright & P. Ayton (Eds.), *Subjective probability* (pp. 239–272). Oxford: Wiley.

Siegler, R. S. (1981). Developmental sequences within and between concepts. *Monographs of the Society for Research in Child Development, 46*, 1–74.

Teglas, 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 of the United States of America, 104*, 19156–19159.

Teglas, 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, 332*, 1054–1058.

Vlamings, P. H., Hare, B., & Call, J. (2010). Reaching around barriers: The performance of the great apes and 3–5-year-old children. *Animal Cognition, 13*, 273–285.

Wellman, H. M., & Liu, D. (2004). Scaling of theory of mind tasks. *Child Development, 75*, 523–541.

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