The Development of Nonsymbolic Probability Judgments in Children

Shaun O’Grady and Fei Xu
University of California, Berkeley

Two experiments were designed to investigate the developmental trajectory of children’s probability approximation abilities. In Experiment 1, results revealed 6- and 7-year-old children’s ($N = 48$) probability judgments improve with age and become more accurate as the distance between two ratios increases. Experiment 2 replicated these findings with 7- to 12-year-old children ($N = 130$) while also accounting for the effect of the size and the perceived numerosity of target objects. Older children’s performance suggested the correct use of proportions for estimating probability; but in some cases, children relied on heuristic shortcuts. These results suggest that children’s nonsymbolic probability judgments show a clear distance effect and that the acuity of probability estimations increases with age.

Children experience a great deal of probabilistic data in everyday life, and both developmental psychologists and educators have found probabilistic reasoning to be a fertile domain for understanding the development of numerical cognition. Throughout development, children encounter a wealth of numerical and non-numerical data, and must integrate these data to make rapid judgments, often based on limited information. Understanding how children leverage their intuitive understanding of number and probability to make decisions in a complex world can provide insights that are relevant to a broad range of fields from perception and decision making to formal mathematics education.

Probabilistic reasoning refers to a broad range of abilities related to uncertainty such as understanding randomness, appropriately analyzing sample spaces, reasoning about correlation, and formally quantifying probability (see Bryant & Nunes, 2012, for a thorough and insightful discussion). This is a broad literature with numerous unanswered research questions. In this article, we focus primarily on children’s estimation of the probability of discrete events, and we aim to chart the developmental trajectory of these abilities. We begin by briefly reviewing the relevant literatures on the development of probabilistic and proportional reasoning abilities as well as the approximate number system (ANS). We then present the results of two experiments designed to investigate the influence of numerical and non-numerical stimulus features on children’s probability judgments and to track the development of the ability to reason about probability based on proportion.

For discrete outcomes, probability is computed as a proportion of target outcomes to all possible outcomes. While a ratio formally describes a relation between two quantities, a proportion is used to assess the equality of two ratios. Although both proportions and ratios can be used to compare probabilities of binary outcomes, comparing ratios can be difficult when the total number of possible outcomes differs between two options. For example, imagine a child is presented with two groups of red and white marbles, and asked to choose the group that is most likely to yield a red marble from a single random draw. Imagine further that one group has 13 red marbles and seven white marbles, whereas the second group has 15 red marbles and nine white marbles. Representing these choices as ratios provides the observer with part:part comparisons, (13:7) and (15:9). Although adults may be adept at computing odds based on ratios, children have difficulty performing these computations and often make their choice based on the group with the larger number of target items (in this case, red marbles) rather than on the proportion of red marbles to total marbles. Proportions facilitate this...
comparison by formalizing the relation between the parts (subsets of outcomes) and the whole (all possible outcomes). In our example, the two proportions would be 13/20 (0.65) and 15/24 (0.625). The first group has a slightly higher chance of yielding a red marble on a single random draw. Thus, the ability to compute proportions can help children accurately judge the equivalence of two probabilities.

For decades, many studies have used the two-alternative forced-choice (2AFC) random draw task to investigate children’s predictions about single and sequential random draws (Falk, Falk, & Levin, 1980; Falk, Yudilevich-Assouline, & Elstein, 2012; Piaget & Inhelder, 1975; Siegler, Strauss, & Levin, 1981; Yost, Siegel, & Andrews, 1962). In this task, children are typically presented with two groups of multiple objects and asked to select the group with the best chance of getting a preferred object. Recently, Falk et al. (2012) conducted a comprehensive study of probabilistic decision making strategies using the 2AFC random draw task with 6- to 12-year-old children. Findings from this study revealed that young children often choose the group with the greatest number of target objects regardless of the total number of objects until around 8 years of age when children begin to attend to the whole set of possible outcomes rather than simply the number of target outcomes. These findings indicate that younger children have difficulty reasoning about probability based on proportion: instead of relating a part (a subset of outcomes) to the whole (all possible outcomes) for each choice and choosing the group with the larger proportion of target outcomes, children merely compare the number of target outcomes in each choice and choose the group with more target outcomes.

Much like the research on probabilistic reasoning, research on proportional reasoning has also shown that children’s ability to reason about proportion greatly improves over the school-age years (Mix, Levine, & Huttenlocher, 1999; Möhring, Newcombe, & Frick, 2015; Singer-Freeman & Goswami, 2001; Spinillo & Bryant, 1999). Many of these studies have deployed the proportional match to sample task in which a child is first presented with a proportional “target” stimuli then presented with several similar proportions from which they should select the item that matches the proportions of the target stimulus. These methods often compare children’s choices when they are presented with proportions in a discrete format (i.e., discrete units of juice and water) to their choices when presented with the same proportions in a continuous format (i.e., portions of juice and water that do not have discrete units). Using this method, researchers have reported a common error in which children choose the item based on matching parts rather than matching proportions (Boyer & Levine, 2012; Boyer, Levine, & Huttenlocher, 2008). This error is similar to the types of incorrect choices made by children in probabilistic reasoning studies discussed earlier (Falk et al., 2012; Piaget & Inhelder, 1975).

In the proportional reasoning literature, research has shown that these errors are most often observed when children are presented with stimuli containing discrete, countable parts (Boyer & Levine, 2012, 2015; Boyer et al., 2008; Hurst & Cordes, 2018; Jeong, Levine, & Huttenlocher, 2007). These findings suggest that young children’s proportional judgments are influenced by their knowledge of whole numbers (Mix et al., 1999; Sophian, 2000; Sophian & Wood, 1997). Based on the findings from the proportional reasoning literature, children can make accurate proportional matches when presented with proportions in a continuous format, but they show a bias toward comparisons of parts when they are presented with proportions in a discrete format. In this article, we seek to chart the developmental trajectory of the ability to compute the probability of discrete events based on proportion.

Humans have remarkable abilities for reasoning about numerical magnitude (Dehaene, 2011; Feigenson, Dehaene, & Spelke, 2004). Our ability to rapidly form accurate approximations of numerical magnitude is often referred to as the ANS and can be observed within the first year of life (Dehaene, Dehaene-Lambertz, & Cohen, 1998; Izard, Sann, Spelke, & Streri, 2009; Lipton & Spelke, 2003; Wood & Spelke, 2005; Xu, 2003; Xu & Spelke, 2000; Xu, Spelke, & Goddard, 2005). In addition to number discrimination, infants form expectations about approximate addition and subtraction (Chiang & Wynn, 2000; McCrink & Wynn, 2004) and can even discriminate ratios (McCrink & Wynn, 2007). Furthermore, young children’s performance in non-symbolic multiplication and division tasks (McCrink & Spelke, 2010, 2016) suggests that ANS representations play a role in arithmetical reasoning even when children have not been formally trained to use algorithms for symbolic multiplication and division.

Decades of research on numerical processing has shown that both humans and nonhuman animals are capable of forming abstract representations of number (Dehaene et al., 1998; Moyer & Landauer,
that young children are capable of making rapid and accurate approximations of probability based on proportions. From this hypothesis, we make three predictions: First, children’s probabilistic discrimination abilities will demonstrate ratio dependence (i.e., as two proportions move further apart on the mental number line, they will become easier to discriminate, also known as the distance effect). Second, the ability to discriminate probabilities will improve with age since both proportional reasoning and number approximation acuity improves with age. Third, based on previous research on probability judgments and proportional reasoning, we predict that probability discrimination will be influenced by the same non-numerical features known to influence perceived numerosity (e.g., size of the objects). Furthermore, we investigate whether children’s nonsymbolic probability estimates will be influenced by the same heuristic decision rules reported in previous research that required counting (Falk et al., 2012); that is, whether children sometimes use only the number of target objects when estimating probability as opposed to the correct proportion strategy.

**Experiment 1: Method**

**Participants**

Sixty 6- to 7-year-old children were recruited from local public schools and museums in the San Francisco Bay area. According to the National Center for Educational Statistics, (NCES [National Center for Education Statistics], 2018), the schools in which we conducted the current series of experiments serve children from a range of racial and ethnic backgrounds (School A: 14% Asian, 5% Black, 11% Latinx, 1% Native Hawaiian, 57% White & 12% Mixed race/ethnicity; School B: 28% Asian, 8% Black, 16% Latinx, 38% White & 10% Mixed race/ethnicity). Although we did not collect data on socioeconomic status, we conduct our experiments at local museums on free admission days in order to recruit families from a range of socioeconomic backgrounds. According to data from the United States Census Bureau, the median incomes of the three communities in which data were collected are $70,393 per year, $92,670 per year, and $140,640 per year (U.S. Census Bureau, 2018), indicating that children in the current sample came from middle to upper-middle class households.

A total of 12 children (10 six-year-olds and 2 seven-year-olds) were excluded because they did not pass practice trials meant to ensure that
participants understood the task. The remaining sample of participants consisted of 24 six-year-olds (N = 24; M_age = 6.28; SD = 0.30; 19 female) and 24 seven-year-olds (N = 24; M_age = 7.62; SD = 0.36; 20 female). Target sample size (N = 48) was determined based on previous research with similar tasks (Fazio et al., 2014; Halberda, Mazzocco, & Feigenson, 2008) as well as the additional constraint of ensuring that only children who passed practice trials were included in the final sample. Because Halberda et al. (2008) investigated age-related differences in a simple dot approximation task with 16 children in each of five age groups and Fazio et al. (2014) collected data for a sample of 53 twelve-year-olds in a ratio comparison task but did not seek to investigate age-related differences, we decided to split the difference and test two age groups with 24 children each.

Material
The images for the task were rendered using Blender 2.72, 3D animation software (http://www.blender.org/). Each image contained two groups of red and white marbles divided by a black partition. Because the goal of this experiment was to investigate the psychophysical properties of probability judgments, we created images with a wide range of proportions. The ratio of the proportions presented in each image ranged from 1.1 (55% vs. 50%) and 14 (70% vs. 5%). Table 1 contains the ratios of the proportions used in Experiment 1. For each ratio of proportion, two trial types were created. The total equal trials contained the same total number of marbles on each side of the partition, whereas the target equal trials contained the same number of target color marbles on each side with the “losing” group containing more nontarget marbles. In total, there were 100 marbles in each group in the total equal trials, whereas target equal trials contained one group with 100 marbles matching the “losing” proportion and another group containing an equal number of target color marbles and enough nontarget color marbles to match the “winning” proportion. Importantly, the difference in the total amount of marbles created a large contrast between the field areas of the “winning” and “losing” groups. In order to reduce the chances that this contrast could cue participants to choose the group with the smallest field area we also created an additional set of “foil” trials in which the “winning” group contained more marbles and thus a larger field area than the “losing” group. Figure 1 contains a visual schematic of the procedure with an example of each trial types as well as an example foil trial image.

Procedure
After their parents signed a written consent form approved by the University of California Berkeley Committee for the Protection of Human Subjects children were asked to provide verbal assent to participate in the study. Children were then seated in front of a MacBook Pro laptop (OSX; Screen resolution 1,280 × 800) and were told that they were going to play a game in which they would help Big Bird collect marbles. Half of the children were instructed to collect red marbles and the other half were asked to collect white marbles.

An experimenter explained that Big Bird could not see the contents of the bags of marbles and that he would take a single marble randomly from the bag that the child chooses. The child was then reminded that Big Bird preferred either red or white marbles and that they should choose the group that was best for getting a marble of that color. The experimenter then told the children that one choice was always better than the other and that some of the trials might seem easy, whereas others may be more difficult. Furthermore, if they were uncertain about which group to choose, they

Table 1
Proportions Presented in Each Trial of Experiment 1

<table>
<thead>
<tr>
<th>Proportion Group 1</th>
<th>Proportion Group 2</th>
<th>Ratio of proportions</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.55</td>
<td>0.50</td>
<td>1.10</td>
</tr>
<tr>
<td>0.70</td>
<td>0.60</td>
<td>1.17</td>
</tr>
<tr>
<td>0.55</td>
<td>0.45</td>
<td>1.22</td>
</tr>
<tr>
<td>0.80</td>
<td>0.60</td>
<td>1.33</td>
</tr>
<tr>
<td>0.80</td>
<td>0.55</td>
<td>1.45</td>
</tr>
<tr>
<td>0.60</td>
<td>0.40</td>
<td>1.50</td>
</tr>
<tr>
<td>0.70</td>
<td>0.40</td>
<td>1.75</td>
</tr>
<tr>
<td>0.55</td>
<td>0.30</td>
<td>1.83</td>
</tr>
<tr>
<td>0.60</td>
<td>0.30</td>
<td>2.00</td>
</tr>
<tr>
<td>0.70</td>
<td>0.30</td>
<td>2.33</td>
</tr>
<tr>
<td>0.80</td>
<td>0.30</td>
<td>2.67</td>
</tr>
<tr>
<td>0.75</td>
<td>0.25</td>
<td>3.00</td>
</tr>
<tr>
<td>0.70</td>
<td>0.20</td>
<td>3.50</td>
</tr>
<tr>
<td>0.80</td>
<td>0.20</td>
<td>4.00</td>
</tr>
<tr>
<td>0.90</td>
<td>0.15</td>
<td>6.00</td>
</tr>
<tr>
<td>0.80</td>
<td>0.10</td>
<td>8.00</td>
</tr>
<tr>
<td>0.90</td>
<td>0.10</td>
<td>9.00</td>
</tr>
<tr>
<td>0.50</td>
<td>0.05</td>
<td>10.00</td>
</tr>
<tr>
<td>0.55</td>
<td>0.05</td>
<td>11.00</td>
</tr>
<tr>
<td>0.70</td>
<td>0.05</td>
<td>14.00</td>
</tr>
</tbody>
</table>

Note. Ratios of proportions are rounded to two decimal points.
should try to make their best guess. In order to reduce the influence of age-related differences in formual understanding of the words “probability” and “proportion,” the experimenter never explicitly mentioned the words “probability” or “proportion” during the instructions. Children were then presented with four practice trials with two groups of marbles, one containing all red marbles, whereas the other contained all white marbles. Participants were told that the practice trials were intentionally easy and were meant to teach them how the game worked.

Each participant was presented with 40 test trials and 10 foil trials in one of two semirandomized orders using the psychophysics toolbox (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997) written for the MatLab programming language. Because previous research using a similar design presented images with fewer objects for 1,320 ms (Fazio et al., 2014) we decided to present the images for 1,500 ms to allow our younger participants more time. Following stimulus presentation, participants saw a screen containing the Big Bird character flanked by two bags labeled with a blue “1” and a green “2”. Participants were instructed to press a key marked by a sticker matching the color of the number on the bag from which they wanted Big Bird to draw a marble. Intermission videos in the form of a 30 s animation were used to give children a break during the game and were presented after the 15th, 30th, and 40th trials. Importantly, children did not receive feedback on their choices until the end of the game at which point every child saw the same screen containing 40 white or red marbles and was told “Wow, you did really good! Look how many red/white marbles you got!”. The computer collected both reaction time and participant choice for each trial. Once the participant completed the last trial, they saw a screen containing 40 marbles that matched their target color and were told that these were the marbles that they had collected during the game. A visual schematic of the procedure is presented in Figure 1.

**Results**

Using the binomial exact test, we find that performance on foil trials was above chance for both 6-year-olds (probability of success = .79, 95% CI [.73, .84], \( p < .001 \)) and 7-year-olds (probability of success = .97, 95% CI [.94, .99], \( p < .001 \)) suggesting that children did not learn to merely select the smallest group when presented with groups of different sizes. Foil trials were excluded from the remainder of analyses.

**Reaction Time**

Reaction time data were cleaned for outliers by excluding reaction times that were either greater or less than 3 median absolute deviations from each participant’s median reaction time. Because the median is relatively insensitive to the effects of outliers compared to the mean, this method is thought
to be a superior method for identifying outlying reaction time data (Ley, Ley, Klein, Bernard, & Licata, 2013). Use of this procedure resulted in the exclusion of 198 of the 1,920 total trials (10.31%). In order to report the most accurate representations of the data, all analyses reported in this article were conducted on the data set in which trials in which outlying reaction time were excluded. Exclusion of these data do not change the results for any of the following analyses including general accuracy and statistical modeling. Results of the same analyses conducted on the complete data set are reported in Supporting Information. Comparisons of performance for all included trials revealed that the reaction time for both age groups was significantly faster on the total equal trials (6-year-olds: $M = 1,121.67$ ms, $SD = 923.22$ ms; 7-year-olds: $M = 642.57$ ms, $SD = 508.97$) compared to the target equal trials (6-year-olds: $M = 1,366.98$ ms, $SD = 1,163.47$; $\Delta M = 245.31$, 95% CI $[-386.98, -103.63]$, $t(781.47) = -3.40$, $p = .001$; 7-year-olds: $M = 758.51$ ms, $SD = 547.90$; $\Delta M = 115.94$, 95% CI $[-186.55, -45.33]$, $t(852.71) = -3.22$, $p = .001$).

**General Accuracy**

Children in both age groups performed significantly above chance on both total equal (6-year-olds: probability of success = .78, 95% CI [.74, .82], $p < .001$; 7-year-olds: probability of success = .91, 95% CI [.88, .93], $p < .001$; binomial exact test) and target equal trial types (6-year-olds: probability of success = .67, 95% CI [.62, .71], $p < .001$; 7-year-olds: probability of success = .85, 95% CI [.82, .89], $p < .001$; binomial exact test). Figure 2 presents the average performance by ratio of proportions and trial type for both age groups.

**Statistical Modeling**

Generalized linear models with mixed effects (GLMMs) predicted the participant’s binary response variable from age, trial type, and ratio of proportions while controlling for the random effects of participant identification number. Preliminary analyses revealed no effects of gender, color of target marble, and order of presentation. For both nested and non-nested models, we use Akaike information criterion (AIC) as a method of model selection. AICs are presented alongside the results of chi-square tests of model fit for nested models.

Comparisons of GLMMs revealed that the model with the best fit to the data predicted the participant’s response based on trial type, participant age group, and the ratio of proportions with no interactions (AIC$_{TT + AG + RP}$ = 1,459.66). This model outperformed the null model (AIC$_{Null}$ = 1,491.36; $\chi^2 = 37.69$; $df = 3$; $p < .001$), the models for trial type (AIC$_{TT}$ = 1,472.13; $\chi^2 = 16.47$; $df = 2$; $p < .001$), and ratio of proportions (AIC$_{RP}$ = 1,486.34; $\chi^2 = 30.67$; $df = 2$; $p < .001$), as well as the models based on trial type and age group (AIC$_{TT + AG}$ = 1,463.98; $\chi^2 = 6.31$; $df = 1$; $p = .01$) and the interaction of trial type and age group (AIC$_{TT × AG}$ = 1,465.86). Furthermore, the models that accounted for the interaction between age and ratio of proportions (AIC$_{TT + AG × RP}$ = 1,461.12; $\chi^2 = 0.55$; $df = 1$; $p = .46$), trial type and ratio of proportions (AIC$_{TT × AG + RP}$ = 1,461.37; $\chi^2 = 0.29$; $df = 1$; $p = .59$), and the three-way interaction between trial type, age, and ratio of proportions (AIC$_{TT × AG × RP}$ = 1,466.07; $\chi^2 = 1.59$; $df = 4$; $p = .81$) were not significantly different from the model without interactions. Importantly, these models have a greater number of parameters, yet they yield relatively inconsequential improvements.

![Figure 2. Average performance by ratio of proportions and trial type for both 6-year-olds (A) and 7-year-olds (B). Error bars indicate bootstrapped 95% CIs.](image-url)
in model fits. In this case, the simpler model is preferred because it explains the same amount of variance with fewer parameters.

The preferred model predicts the participant’s binary response based on trial type, age group, and the ratio of proportions of the presented image \( (\beta_{\text{Intercept}} = 0.75; \ SE = .26; \ 95\% \ CI \ [0.24, \ 1.26]). \) Inspection of the exponentiated model coefficients revealed that \( \text{total equal} \) trials led to an 85% increase in the odds of obtaining a correct answer \( (\beta_{TT} = 0.62; \ SE = .14; \ 95\% \ CI \ [0.35, \ 0.89]). \) The main effect of age indicated that 7-year-olds performed better than 6-year-olds with the odds of a correct response increasing by a factor of 3.25 for 7-year-olds compared to 6-year-olds \( (\beta_{\text{AG}} = 1.18; \ SE = .35; \ 95\% \ CI \ [0.49, \ 1.87]). \) Last, we report a main effect of the ratio of proportions with a unit increase in ratio of proportions leading to a 5% increase in the odds of a correct response \( (\beta_{\text{RP}} = 0.05; \ SE = .02; \ 95\% \ CI \ [0.01, \ 0.09]). \) Analyses of reaction times yielded similar results details of which can be viewed in Supporting Information.

Discussion

In Experiment 1, our results showed that 6- and 7-year-old children’s nonsymbolic probability judgments were predicted by the ratio of proportions (i.e., the distance effect). As the ratio of proportions of the two distributions becomes larger, performance improved. These findings are consistent with similar studies with adults (O’Grady, Griffiths, & Xu, 2016; O’Grady et al., submitted) and older children (Fazio et al., 2014). Falk et al. (2012) report that children’s probabilistic judgments gradually improve with age and that by the age of 8 children are capable of using the correct proportional strategy. Results from Experiment 1 support these findings. Although 7-year-old children performed better than 6-year-old children, both age groups performed worse on \( \text{target equal} \) trials compared to \( \text{total equal} \) trials, indicating that on some trials children may have focused on either the number of target objects or the number of nontarget objects without relating the two quantities.

Although the results from Experiment 1 provide novel insight into how young children approximate and reason about binary probabilities, three important design features limit the strength of our findings. First, the images in Experiment 1 consisted of marbles neatly arranged into orderly rows and columns which may have helped some children more accurately approximate the number of marbles in each group. Second, the use of \( \text{target equal} \) trials makes it difficult to assess whether participants were focusing on the number of target objects or the number of nontarget objects. Last, all the marbles in Experiment 1 were the same size yet much of the research using dot approximation tasks has indicated that number approximation is influenced by non-numerical stimulus features such as size and sparsity (Allik, Tuulmets, & Vos, 1991; DeWind, Adams, Platt, & Brannon, 2015; Starr, DeWind, & Brannon, 2017). In order to address these concerns, we designed new stimuli consisting of (a) smaller numbers of marbles randomly positioned on the screen, (b) trials in which the group with a larger proportion of target color marbles contained fewer marbles of the target color than the group with the smaller proportion of target color marbles, (c) trials in which the target marbles in the “losing” distributions were larger than the target marbles in the “winning” distribution. Because we expected each of these changes to increase the difficulty of the task and performance of the 6-year-olds in Experiment 1 was already relatively low, we decided to test older children for Experiment 2.

Experiment 2: Method

Participants

One hundred and forty-two children between the ages of 7 and 12 were recruited from local schools and children’s museums from the San Francisco Bay area. Twelve of these children were excluded from our analyses: eight children were excluded due to experimenter error, three children did not pass the practice trials, and one child’s parent interfered in the study by coaching their child to choose the group with a larger proportion of target marbles. As with Experiment 1 our target sample size was determined based on previous research (Fazio et al., 2014). However, because our target age range was much larger (7- to 12-year-olds), data collection continued according to a stopping rule requiring a minimum of 40 participants in each of three age groups (7- to 8-year-olds, 9- to 10-year-olds, and 11- to 12-year-olds). The final sample consisted of forty 7- and 8-year-olds \( (N = 40; \ M_{\text{age}} = 7.96; \ SD = 0.53; \ 24 \text{ female}), \) fifty 9- and 10-year-olds \( (N = 50; \ M_{\text{age}} = 10.04; \ SD = 0.50; \ 20 \text{ female}), \) and forty 11- and 12-year-olds \( (N = 40; \ M_{\text{age}} = 11.75; \ SD = 0.64; \ 18 \text{ female}). \) Data collection was conducted in the same schools and communities reported in Experiment 1.
Material

As mentioned above, the orderly arrangement of marbles in Experiment 1 may have helped children approximate the number of marbles in each group. In order to prevent this, the location of each marble was randomly generated for each image using Blender 2.72. Due to the ceiling levels of performance for high ratios of proportions in Experiment 1 we decided to include more trials with lower ratios of proportions, ranging from 1.1 (55% vs. 50%) to 9.5 (95% vs. 10%). We also decided to include ratios of two proportions that were both below chance (i.e., 40% to 15%). Table 2 presents the proportions of marbles in each group alongside the ratios of proportions used in Experiment 2. For each ratio of proportions, three trial types were created: *total equal* trials in which each distribution had the same total number of marbles; *area-anticorrelated* trials in which the sizes of the marbles were manipulated such that the total area covered by the target marbles in the “losing” group was larger than the total area covered by the target marbles in the “winning” group. Importantly, *area-anticorrelated* trials included groups of marbles with an equal total amount of marbles similar to *total equal* trials. Finally, *number versus proportion* trials in which the distribution with the lower proportion of target marbles contained a larger number of target marbles. A total of 264 images were rendered and then divided equally into four conditions based on target color and the order of presentation of the images. Each participant viewed 66 images presented in one of four conditions (Red, Order 1; Red, Order 2; White, Order 1; and White, Order 2). Importantly, the order of the images was pseudorandomized such that there were no more than 3 consecutive trials in which the “correct” choice was on the same side of the screen.

### Procedure

After parental guardians provided written consent for their children to participate in the study, children were asked to provide verbal assent and were then seated in front of a MacBook Pro laptop (OSX; Screen resolution 1,280 × 800). Participants were told that they were going to play a game in which they would collect red or white marbles depending on the condition to which they were assigned. Children were shown two boxes and told that they would see two groups of marbles on two trays on the screen. The group of marbles on the left side of the screen were poured into the box on the left and the group on the right side of the screen were poured into the box on the right side of the screen. The boxes would then be shaken up so that they could not infer the positions of the marbles based on their locations on the viewing trays. They were then asked to select the box that they thought was best for collecting their target color marble. After the instructions phase participants played four practice trials in order to ensure that they understood the game. Once the practice trials were complete, participants were presented with 66 semirandomized test trials in which they were able to view the images for 1,500 ms before making their selection. As with Experiment 1 short intermission videos were played after the 15th, 30th, and 45th trials, children were not given any feedback about their decisions, and the experimenter never mentioned the words “probability” or “proportion.” Figure 3 provides a visual schematic of the procedure.

### Results

#### Reaction Time

The same procedure employed in Experiment 1 for cleaning outlying reaction time resulted in the exclusion of 1,383 of 8,580 trials (16.12%). As with

Table 2

<table>
<thead>
<tr>
<th>Proportion Group 1</th>
<th>Proportion Group 2</th>
<th>Ratio of proportions</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.55</td>
<td>0.50</td>
<td>1.10</td>
</tr>
<tr>
<td>0.50</td>
<td>0.45</td>
<td>1.11</td>
</tr>
<tr>
<td>0.45</td>
<td>0.40</td>
<td>1.12</td>
</tr>
<tr>
<td>0.55</td>
<td>0.45</td>
<td>1.22</td>
</tr>
<tr>
<td>0.80</td>
<td>0.60</td>
<td>1.33</td>
</tr>
<tr>
<td>0.80</td>
<td>0.55</td>
<td>1.45</td>
</tr>
<tr>
<td>0.75</td>
<td>0.50</td>
<td>1.50</td>
</tr>
<tr>
<td>0.60</td>
<td>0.40</td>
<td>1.50</td>
</tr>
<tr>
<td>0.95</td>
<td>0.55</td>
<td>1.73</td>
</tr>
<tr>
<td>0.95</td>
<td>0.50</td>
<td>1.90</td>
</tr>
<tr>
<td>0.50</td>
<td>0.25</td>
<td>2.00</td>
</tr>
<tr>
<td>0.40</td>
<td>0.15</td>
<td>2.67</td>
</tr>
<tr>
<td>0.60</td>
<td>0.20</td>
<td>3.00</td>
</tr>
<tr>
<td>0.50</td>
<td>0.15</td>
<td>3.33</td>
</tr>
<tr>
<td>0.80</td>
<td>0.20</td>
<td>4.00</td>
</tr>
<tr>
<td>0.40</td>
<td>0.10</td>
<td>4.00</td>
</tr>
<tr>
<td>0.75</td>
<td>0.15</td>
<td>5.00</td>
</tr>
<tr>
<td>0.95</td>
<td>0.15</td>
<td>6.33</td>
</tr>
<tr>
<td>0.80</td>
<td>0.10</td>
<td>8.00</td>
</tr>
<tr>
<td>0.85</td>
<td>0.10</td>
<td>8.50</td>
</tr>
<tr>
<td>0.90</td>
<td>0.10</td>
<td>9.00</td>
</tr>
<tr>
<td>0.95</td>
<td>0.10</td>
<td>9.50</td>
</tr>
</tbody>
</table>

*Note.* Ratios of proportions are rounded to two decimal points.
Experiment 1, exclusion of trials with outlying reaction time does not change the reported results. A report of the results of the same analyses for the complete data set for Experiment 2 are included in Supporting Information. Of the included trials, reaction times were significantly lower on total equal trials ($M = 815.29$ ms, $SD = 466.83$) and number versus proportion trials ($M = 819.06$ ms, $SD = 441.59$) compared to area-anticorrelated trials ($M = 888.15$ ms, $SD = 480.37$; total equal: $ΔM = 72.86$, 95% CI $[99.77, 45.95]$, $t(4,758.99) = 5.31$, $p < .001$; number versus proportion: $ΔM = 69.09$, 95% CI $[95.08, 43.11]$, $t(4,805.06) = 5.21$, $p < .001$). The difference between total equal trials and number versus proportion trials did not reach significance ($ΔM = 3.77$, 95% CI $[29.55, 22.00]$, $t(4,741.64) = .29$, $p < .774$).

**General Accuracy**

Results of the binomial exact tests comparing performance against chance revealed that children in all three age groups performed significantly above chance on both total equal trials (8-year-olds: probability of success $= .78$, 95% CI $[.75, .81]$, $p < .001$; 10-year-olds: probability of success $= .87$, 95% CI $[.85, .89]$, $p < .001$; 12-year-olds: probability of success $= .91$, 95% CI $[.88, .93]$, $p < .001$) and area-anticorrelated trials (8-year-olds: probability of success $= .60$, 95% CI $[.57, .64]$, $p < .001$; 10-year-olds: probability of success $= .69$, 95% CI $[.66, .72]$, $p < .001$; 12-year-olds: probability of success $= .61$, 95% CI $[.57, .64]$, $p < .001$). Finally, 12-year-old and 10-year-old children performed significantly better than chance on number versus proportion trials (10-year-olds: probability of success $= .55$, 95% CI $[.52, .58]$, $p < .003$; 12-year-olds: probability of success $= .61$, 95% CI $[.57, .64]$, $p < .001$), whereas 8-year-olds’ performance was not significantly different from chance on number versus proportion trials (8-year-olds: probability of success $= .46$, 95% CI $[.43, .50]$, $p < .056$). Figure 4 presents the proportion of correct responses by ratio of proportions, trial type, and age group.

**Statistical Modeling**

As with Experiment 1, we compared GLMM and used AIC as our method for model selection for non-nested models and chi-square tests for nested models. Results of the model comparisons revealed that the model predicting performance from the three-way interaction between trial type, ratio of proportions, and age group ($AIC_{FullModel} = 7,412.06$) outperformed all other models including the null model ($AIC_{Null} = 8,503.15$; $χ^2 = 1,125.10$; $df = 17$; $p < .001$), the models for trial type ($AIC_{TT} = 7,887.38$; $χ^2 = 505.32$; $df = 15$; $p < .001$), the interaction of trial type and ratio of proportions ($AIC_{RP} = 8,147.82$; $χ^2 = 767.76$; $df = 16$; $p < .001$) as well as more complex models based on trial type and age group ($AIC_{TT + AG} = 7,864.39$; $χ^2 = 478.33$; $df = 13$; $p < .001$), trial type and ratio of proportions ($AIC_{TT + RP} = 7,488.82$; $χ^2 = 104.77$; $df = 14$; $p < .001$), the interaction of trial type and ratio of proportions ($AIC_{TT × RP + AG} = 7,444.83$; $χ^2 = 80.47$; $df = 14$; $p < .001$).

Figure 3. Diagram of the experimental procedure used in Experiment 2. The sample image at the top presents an area-anticorrelated trial, the sample image in the middle presents a total equal trial and the sample image at the bottom presents a number versus proportion trial.
df = 12; p < .001), and the interaction of trial type and age group (AIC TT × AG + RP = 7,444.83; χ² = 68.53; df = 8; p < .001).

Coefficients in logistic regression indicate the change in log odds of a correct response based on changes in experimental and subject variables. It is easiest to express these changes by exponentiating the coefficients to reveal the change in odds and to relate these changes to the baseline group: 8-year-olds’ responses to total equal images (β_intercept = 0.88). Exponentiated model coefficients for main effects of trial type revealed number versus proportion trials led to an 82% reduction in the odds of a correct response (β_{NvP} = -1.73), whereas area-anticorrelated trials only lead to a 52% decrease in the odds (β_{AA} = -0.75). Main effects of age indicated that performance improved with each age group (β_{Age10} = 0.25; β_{Age12} = 0.21). Because the ratio of proportions is a continuous variable, the associated coefficient revealed that a single incremental change in ratio of proportions led to a 15% increase in the odds of a correct response (β_{RP} = 0.14). The only interaction term to reach significance indicated that the effect of ratio of proportions increased for 12-year-olds compared to the younger children (β_{Age12} × RP = 0.14). The full set of fixed effect model coefficients are presented in Table S4 in Supporting Information. Analyses of reaction time data from Experiment 2 revealed a similar pattern of results, the details of which are available in Supporting Information.

Discussion

Experiment 2 replicated the main results of Experiment 1 and extended these findings by including three trial types (i.e., total equal, number vs. proportion, and area-anticorrelated) and three age groups. As in Experiment 1, we found that the performance of children of all three age groups was strongly influenced by the ratio of proportions, converging with results from adults using a very similar methodology (O’Grady et al., submitted) and those of Fazio et al. (2014) with 12-year-old children. By including area-anticorrelated trials, our results revealed that children relied on both numerical and non-numerical stimulus features; that is, children made more errors when the total area of the target marbles was larger in one distribution even when the proportion of target marbles was smaller in that distribution. The model coefficient for area-anticorrelated trials suggests that children of all ages were influenced by this manipulation, indicating that non-numerical stimulus features influence probability judgments. By including number versus proportion trials, our results revealed that children up to 10 often used a formally incorrect strategy in estimating proportions and probability; that is, they used the number of target marbles in a distribution as a proxy for estimating the proportion of target marbles. The only age group that performed above chance level on these trials was the oldest (12-year-olds), and their accuracy was far from perfect (only 60%).

General Discussion

In two experiments, we provide evidence that 6- to 12-year-old children can make rapid and accurate approximations of probability based on proportions. Our findings are consistent with the results of other ratio comparison tasks with adults (O’Grady et al., submitted), 12-year-old children (Fazio et al., 2014) and nonhuman primates (Drucker, Rossa, & Brannon, 2016). More importantly, our results shed new light on the development of proportional and
probabilistic reasoning. In Experiment 1 we report
data demonstrating that 6- and 7-year-olds’ non-
symbolic probability judgments are characterized
by ratio dependence and that the acuity of these
representations improves with age. Experiment 2
replicated these findings and also revealed that
nonsymbolic probability judgments are influenced
by the same numerical and non-numerical stimulus
features which influence perceived numerosity such
as the size of dots in a dot discrimination task. Data
from both experiments also suggest that children
produced similar errors in our nonsymbolic proba-
bility estimation task compared previous research
using the 2AFC random draw task (Falk et al.,
2012) as well as studies on children’s proportional
reasoning (Boyer et al., 2008; Hurst & Cordes,
2018). More specifically, children’s performance was
influenced by the number of target marbles as evi-
denced by their decreased performance on number
versus proportion trials relative to total equal trials.

These findings make three important contribu-
tions to the literature on probabilistic reasoning,
proportional reasoning, and quantitative develop-
ment in general. First, we provide evidence that
children’s probability judgments are characterized
by ratio dependence and even young children can
make accurate judgments about the likelihood of
future events based on proportions. Second, our
current experiments represent the first attempt to
systematically investigate the mental representation
and psychophysical properties of nonsymbolic
probability in the developing human mind. We pro-
vide evidence that school-age children’s probability
estimation is influenced by the size of the objects
being approximated. Third, previous research has
not charted the developmental trajectory of the
mental representation of nonsymbolic probability.
We provide the first evidence that between 6 and
12, children’s ability to estimate probability
improves with age, and they gradually adopt the
correct proportion strategy although they continue
to make errors by sometimes employing heuristic
decision rules.

The results of the current experiments provide
new insights on the role of proportional reasoning
in children’s probability judgments; they also raise
important questions for future research. In the cur-
rent studies, 9- to 12-year-old children performed
above chance in the number versus proportion tri-
als, whereas the 8-year-olds did not. In contrast,
there is some evidence suggesting that both infants
and nonhuman primates can use ratio of propor-
tions in estimating probability (Denison & Xu, 2014;
Rakoczy et al., 2014). In one study, infant
participants observed an experimenter randomly
draw a single lollipop from each of two groups of
preferred and nonpreferred color lollipops (Denison
& Xu, 2014). Infants were more likely to approach
the lollipop drawn from the distribution with a lar-
ger proportion of preferred lollipops even when the
total number of lollipops in both groups varied
such that the group with the lower proportion actu-
ally contained more of the infant’s preferred lol-
lipops. It may be the case that ANS acuity
improves with age, and the current studies used
ratios of proportions that were more difficult than
that of Denison and Xu (2014). However, in Experi-
ment 2 of Denison and Xu (2014), infants performed
above chance when presented a ratio of proportions
of 4 (80% target objects in one distribution vs. 20%
target objects in the other distribution). The current
studies included the same ratio of proportions yet it
is not until about 9 that children succeeded on the
number versus proportion trials. One possible
explanation for this is that older children’s poor
performance on these trials may be due to the
“whole number bias” reported in the education lit-
erate on rational number learning (Ni & Zhou,
2005). In the fraction learning literature, the “whole
number bias” is observed most often when children
choose the larger of two fractions based on the
magnitude of the components of the fractions (i.e.,
by choosing the fraction with the larger numerator
or denominator) rather than selecting the larger
fraction based on the relation between numerator
and denominator. The literature on probability rea-
soning has investigated this same response bias in
the context of probability predictions beginning
with the seminal work of Piaget and Inhelder
(1975), and recent work (Falk et al., 2012) has indi-
cated that this type of response bias constitutes a
strategy that younger children use in 2AFC proba-
bility judgment tasks.

The integrated theory of numerical development
(Siegler, 2016; Siegler, Thompson, & Schneider,
2011) posits that children come to understand
rational numbers through analogy to whole num-ers and evidence from studies on proportional rea-
soning suggests that children overextend their
knowledge of whole numbers when reasoning
about proportions presented as discretized units
rather than continuous quantities (Boyer et al.,
2008; Mix, Huttenlocher, & Levine, 2002). It is pos-
sible that an overreliance on whole number knowl-
dge led to younger children’s incorrect choices on
number versus proportion trials. To explore this pos-
sibility, we are currently developing a modified ver-
sion of our probability discrimination task for use
with much younger, preschool-age children and toddlers. The prediction is that much younger children may succeed in using proportions to estimate probability (consistent with the findings with infants) whereas older, school-age children may adopt the whole number strategy. Indeed, preliminary evidence (O’Grady & Xu, 2018) has shown that school-age children demonstrate a whole number bias when making probability judgment tasks involving both exact and approximate quantities but this bias can be overridden if the child is provided with enough feedback.

Our experiments also raise new questions about the role of magnitude processing in proportional reasoning and probabilistic estimation. The current body of literature suggests two possibilities. One is that the ANS serves as a building block for computing probabilities. According to this account, children first approximate the number of marbles of each type within each group and then use these approximate representations to compute the probabilities. Specifically, probabilities are computed as follows: (number of target objects)/(number of target objects + number of non-target objects). A second possibility suggests that children bypass discrete number approximations altogether and simply approximate ratios using a ratio processing system (RPS). Recent research has provided a wealth of evidence to suggest that ratio processing is fundamental to human numerical cognition (Matthews & Chesney, 2015; Matthews & Lewis, 2017), and thus constitutes a basic building block for learning symbolic fractions (Matthews, Lewis, & Hubbard, 2016). Although we agree that ratio processing is foundational for mathematics learning, it remains unclear whether the RPS and the ANS are two separable systems. Indeed researchers studying early numerical development have recently argued that there exists a general magnitude processing system in the brain, that includes estimations of number (integers, proportions, and probability), duration, and spatial extent (Lourenco, 2016; Mix, Levine, & Newcombe, 2016). Thus, we tend to favor the former claim that children draw on ANS representations for three reasons. First, it remains unclear at the moment whether the RPS exists independently of the ANS. Second, in order to calculate the probability of a discrete event, decision makers must represent discrete outcomes. It is possible that the RPS may be able to compute proportions of discrete elements and this is exactly the type of argument that would support the notion that RPS and ANS are two elements of a more generalized magnitude processing system. Indeed, Jacob, Valentin, and Nieder (2012) have suggested that the ANS may provide one source of input for the RPS. Third, children’s performance on number versus proportion trials in our experiment suggests that number approximations may play an important role in their probability judgments. This claim is clearly speculative based on the current series of studies, but it provides an important avenue for future research and the domain of probabilistic reasoning offers an interesting way to study the relationship between the ANS and RPS.

Last, the current studies are also limited in the types of probabilistic reasoning they address. Based on the findings in the proportional reasoning literature (Boyer et al., 2008), a natural extension of the current work is to investigate whether children rely on the same heuristic decision rules we find for discrete probability (i.e., marbles drawn from a container) when making judgments about continuous probability (i.e., spinner tasks). Furthermore, the current set of experiments focus exclusively on simultaneously presented visual information, thus the role of number and ratio approximation in judgments about sequentially presented probability problems, similar to the methods reported in Boyer (2007), cannot be addressed by the current findings. Future work will investigate whether children and adults rely on ratio processing and integer approximation when tracking and computing the probability of sequentially presented stimuli.

These findings indicate that children can make rapid estimations about the probability of discrete outcomes. Furthermore, we have shown that these representations share some common features with perceptual systems for processing numerical magnitude. By linking the developmental literatures on the ANS and probabilistic reasoning we have demonstrated children’s intuitive ability to estimate probability is surprisingly accurate. Although our results are perhaps most relevant to researchers and educators studying the development of numerical cognition and quantitative development, they may also inform research from a variety of subfields in developmental psychology, such as the development of decision-making strategies and scientific reasoning.

All methods, analyses, and deidentified data are available on the Open Science Framework (https://osf.io/48sgv/).

References


**Supporting Information**

Additional supporting information may be found in the online version of this article at the publisher’s website:

**Appendix S1.** Reaction Time Analyses and Additional Information.