Whole number bias in children's probability judgments.

Shaun O'Grady
shaun.ogrady@berkeley.edu
Department of Psychology
University of California, Berkeley

Fei Xu
fei_xu@berkeley.edu
Department of Psychology
University of California, Berkeley

Abstract

Simple probability judgments pervade human experience. Decades of research have revealed a pattern of heuristic errors in simple random draw predictions of both children and adults. Participants often make their choice based on the magnitude of the target or the non-target set without relating the two quantities. In a series of experiments, we demonstrate that this bias is robust in both timed and untimed tasks (Experiment 1) and may be overcome when the child is given the adequate amount and type of feedback (Experiment 2).

Keywords: Probabilistic reasoning; proportional reasoning; cognitive development; numerical cognition

Introduction

Probabilistic reasoning provides the developing mind with a powerful domain-general tool for making use of the highly variable data encountered throughout life. Children in the US learn the formal principles of probability theory in school around the age of 12, yet research on children's understanding of probability indicates children develop intuitions about uncertain outcomes at much younger ages (Denison & Xu, 2014; Falk, Yudilevich-Assouline, & Elstein, 2012; O'Grady & Xu, submitted; Teglas, Girotto, Gonzalez, & Bonatti, 2007; Xu & Garcia, 2008). What is the relation between a child's intuitive understanding of probability and their formal understanding?

Children's ability to quantify probability has been studied for decades. Piaget & Inhelder (1975) were the first to use the 2-alternative forced choice (2AFC) random draw task to assess children's ability to use quantity information. In this task, the child is asked to choose between two distributions with varying amounts of different color tokens. Decades of research on the topic have led to methodological and procedural refinements (Chapman, 1975; Falk et al., 2012; Fischbein, Pampu, & Mnzat, 1970; Yost, Siegel, & Andrews, 1962) with the most recent of these reports offering a more accurate assessment of probabilistic decision making.

Falk et al. (2012) devised a strategy assessment task involving 24 binary random draw comparisons to study children's use of rule-based reasoning in the random draw task. Falk et al. (2012) distinguished between four possible strategies that children use in the self-paced task. Children are thought to transition through the four strategies as their understanding of probability becomes more sophisticated. At first, children focus on one-dimension of the problem such as the target or non-target events. One-dimensional strategies include choosing the distribution with a greater number of winning marbles ('greater win') as well as choosing the distribution with the smaller number of losing marbles ('lower loss'). Eventually they learn to integrate the two-dimensions into more complex strategies. Two dimensional strategies include choosing the bin with the greater difference between winning and losing beads ('greater difference') as well as the correct proportional strategy ('correct proportion', i.e., number of winning beads out of the total of winning and losing beads). Importantly, Falk et al. (2012) found that by age 8, about half of the children in their sample demonstrated the ability to use the correct proportional strategy and the proportion of children using this strategy increased with age (from 4-11).

Decades of research on proportional reasoning and rational number processing more broadly have shown that young children (Boyer, Levine, & Huttenlocher, 2008; McCrink & Spelke, 2016; Mix, Levine, & Huttenlocher, 1999; Sophian & Wood, 1997) and even infants (Denison & Xu, 2014; McCrink & Wynn, 2007) are capable of accurately representing proportions. Yet, evidence with older children suggests their probabilistic decisions are sometimes biased toward distributions with a greater number of target marbles (O'Grady & Xu, submitted). In this study, we used the 2AFC design with large numbers of marbles presented for a short amount of time, forcing the participant to rely on approximate representations of number. Results revealed that children and adults show a 'whole number bias' for non-symbolic ratio comparison tasks (O'Grady, Griffiths, & Xu, 2016) and this bias tends to decline with age (O'Grady & Xu, submitted). However, it is possible that this 'whole number bias' could be an artifact of the timed nature of the task.

In Experiment 1, we investigate whether children's response bias is an artifact of the timed nature of the approximation task or whether it indicates an inaccurate understanding of the proportional nature of probability. We hypothesize that children use the same heuristic-based decision strategies when reasoning about both exact and approximate non-symbolic probabilities. Based on this hypothesis we predict children will make similar errors in a timed 'probability approximation task' and an 'exact numerical value probability task' involving comparisons of proportions. In Experiment 1 we test this by asking children to perform the probability approximation task as well as a computerized version of the strategy assessment task used by Falk et al. (2012).

Experiment 1 Methods

Participants

Fifty children between the ages of 7 and 10 were recruited from museums, schools and homes in the San Francisco Bay area (N = 50; 9 7-year-olds, Mean age = 7.59, SD = 0.34; 11 8-year-olds, Mean age = 8.3, SD = 0.35; 14 9-year-olds, Mean age = 9.45, SD = 0.34; 16 10-year-olds, Mean age = 10.32, SD = 0.27). Parents of the children provided written informed consent in accordance with regulations established by the UC Berkeley Committee for the Protection of Human Subjects. Although Falk et al. (2012) recruited children ages 4-12, we decided to focus our recruiting efforts on the older children as the older children were more likely to understand the proportional nature of probability.

Material

We deployed 2 different 2AFC tasks using the psychophysics toolbox written for the MatLab programming language. In both tasks, we present children with images depicting two different groups of marbles separated by a black partition. Each group contained marbles with two different colors and participants were asked to choose one of the two groups in order to draw a marble of a certain color. In both games, images were created using Blender (Version 2.78). Figure 1 presents example images for each trial type used in both tasks.

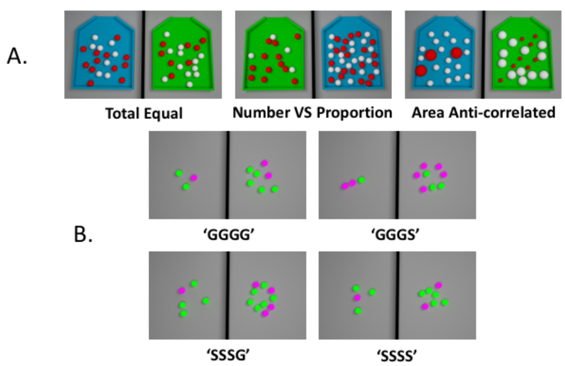

Figure 1: Example images for trial types in both probabilty approximation task (A) as well as the exact numerical value task (B). Red is the target color marble for the imgaes in (A) and green is the target color marble in (B).

Probability Approximation Task Images contained two groups of marbles which varied on the ratio of proportions of target marbles. Importantly, three different trial types were used in this task: (1) 'total equal' trials contained the same total number of marbles in each group, (2) 'number vs proportion' trials contained a larger number of target marbles in the group with a lower proportion of target marbles, and (3) 'area-anticorrelated' trials contained marbles which varied in size such that the surface area of target color marbles was a greater proportion of the total surface area in the 'losing' distribution and a smaller proportion of total surface area in the 'winning distribution'. If children rely on faulty heuristic rules such as 'pick the group with the largest number of target marbles' their performance would be significantly lower on 'number vs proportion' trials compared to both 'total equal' and 'area-anticorrelated' trials. Importantly, each image was presented for 1500ms to prevent children from counting.

Exact Numerical Value Probability Task Images consisted of two small groups of green and purple marbles and trials fell into four categories which were used by Falk et al. (2012) to differentiate the 4 different strategies children can use. Following Falk et al. (2012) each trial was internally labeled with the number of green and purple marbles as well as the trial type designators 'GGGG', 'GGGS', 'SSSG', and 'SSSS'. Briefly, the placement of letters represents the dimension of comparison and the letter itself relates the 'winning' to the 'loosing' choice. For example, on 'GGGG' trials, the 'winning' distribution has a greater number of target marbles (1st G), a greater number of non-target marbles (2nd G), a greater sum of both target and non-target (3rd G) and a greater difference between target and non-target marbles (4th G). For comparison, trials marked with 'SSSS' consist of distributions in which the 'winning' choice contains fewer target, fewer non-target, a smaller total, and smaller difference of marbles than the 'losing' choice. Children were instructed to count all of the marbles in each of the groups on the screen before making a decision. This is in contrast to the 'probability approximation task' in which children are prevented from counting by a brief stimulus presentation time and a large amount of marbles in each group.

Procedure

Children were seated approximately 60 cm away from a MacBook Pro laptop (OSX; Screen resolution 1280 x 800) on which they viewed the images for each task. For logistical reasons, 40 of the 50 participants performed the 'probability approximation task' first and an additional 10 participants were recruited for the reverse order once data collection for O'Grady & Xu (submitted) was complete. In the 'probability approximation task' children were instructed to choose the bin they thought was best of getting a target color marble. During the instructions for the 'exact numerical value probability task', children were given the same instructions except they were prompted to count all of the marbles in each group before making a choice and were informed that they could take as long as they needed to make a decision. Children completed 24 trials in this manner with a brief intermission after the first 12 trials.

Experiment 1 Results

We first calculate each participant's strategy based on their responses to the 24 trials of the 'exact numerical value probability task'. Results revealed that performance was significantly positively correlated with age (Pearson's $r = .31$, 95% CI $[.04, .54]$, $t(48) = 2.28$, $p = .027$) indicating that children rely on more complicated and accurate strategies as they get older. Figure 2 presents the proportion of children using each of the 4 strategies by age group. Next, we investigated the relation between children's strategy as measured by the 'exact numerical value probability task' and their performance on different trial types in the 'approximate probability task'. Ten participants did not complete the 'approximate probability task' leaving a total of 40 participants with which to compare the two tasks. Within this subsample, we find that (24) children (60%) used the 'greater win', none of the children used the 'lower loss' strategy, (6) children (15%) used the 'greater difference' strategy, and (10) children (25%) used the formally correct proportional strategy.

Children's strategies in the 'exact numerical value' task were correlated with their performance in the 'probability approximation' task (Pearson's $r = .48$, 95% CI $[.19, .69]$, $t(38) = 3.34$, $p = .002$). Furthermore, this correlation was driven by performance on 'number vs proportion' trials ($r = .66$, 95% CI $[.44, .81]$, $t(38) = 5.47$, $p < .001$) in the probability approximation task as children's strategies did not correlate with performance on either 'total equal' trials ($r = .10$, 95% CI $[-.22, .40]$, $t(38) = 0.64$, $p = .526$) or 'area anti-correlated' trials ($r = .11$, 95% CI $[-.21, .41]$, $t(38) = 0.69$, $p = .495$). Figure 3 presents the proportion of correct choices by ratio of proportions for children using the 1-dimensional (A) and 2-dimensional strategies (B). The diamond symbols on the far-right side of the graph indicate the proportion of correct responses for the easiest ratio of proportion comparisons and those on the left side represent the most difficult comparisons. Note that the children using the 2-dimensional strategies perform at greater than chance levels on many 'number vs proportion' trials while children employing the 1-dimensional strategy are largely at or below chance for these same trials as indicated by the red diamonds.

Experiment 1 Discussion

In the 'exact numerical value probability task', children were instructed to count the marbles in each group and in the 'probability approximation' task they were prevented from counting yet performance on the two tasks was correlated. More specifically, children relying on 2-dimensional strategies outperformed children using 1-dimensional strategies on the 'number vs proportion' trials of the probability approximation task. These results suggest that children used the similar decision making strategies when they knew the exact numerical values of small numbers as well as when they approximated large numbers of marbles presented for 1.5 seconds. However, since the two tasks were not properly counterbalanced, it is impossible to completely rule out order effects.

While it is clear that children's strategy use improves with age it is unclear how this learning process unfolds. In Experiment 2, we investigate the influence of feedback on children's strategy use by first assessing children's strategies and then we presenting them with a series of 2AFC random draw task trials during which children are given feedback. Finally, we present test trials designed to investigate whether children changed their strategy. Based on previous research (Falk et al., 2012), we hypothesize that young children are capable of learning to use the correct strategy but only when provided with examples that do not fit their incorrect understanding. We predict that children can learn to make correct choices in the 2AFC random draw task if they are presented with trials that conflict with their strategy.

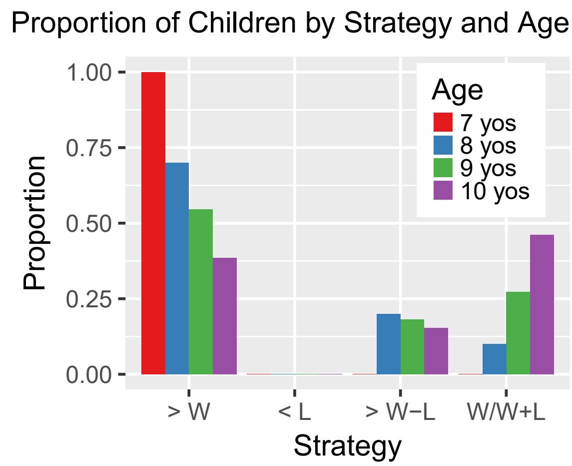

Figure 2: Proportion of children using each strategy by age group. The strategies listed along the x-axis are 'greater win' (> W), 'lower loss' (< L), 'greater difference' (> W - L) and 'higher proportion' (> W/W+L).'

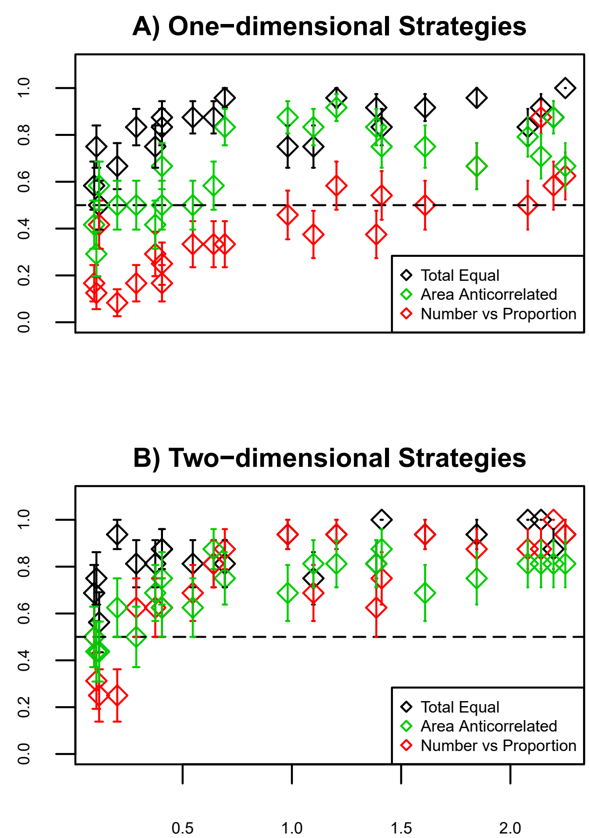

Figure 3: Performance on the 'probability approximation task' by ratio of proportions and trial type. A) Children using one-dimensional strategies. B) Children using two-dimensional strategies' task. The x-axis plots the log of the ratio of proportions for trials in the 'probability approximation task'.'

Experiment 2 Methods

Participants

Fifty-seven children between the ages of 6 and 11 were recruited from museums, schools, and homes in the San Francisco Bay area. Data from ten children were excluded from this sample. One child decided stopped the game early, three children were coached by their parents, and four children were excluded due to equipment malfunction or experimenter error. An additional two children were excluded because their average reaction time on the ***assessment phase*** was lower than 3 seconds and thus did not have enough time to count the marbles as instructed. Our current sample consisted of N = 47 children (5 6-year-olds, Mean age = 6.52, SD = 0.13; 5 7-year-olds, Mean age = 7.5, SD = 0.29; 7 8-year-olds, Mean age = 8.53, SD = 0.25; 17 9-year-olds, Mean age = 9.5, SD = 0.32; 9 10-year-olds, Mean age = 10.5, SD = 0.29;4 11-year-olds, Mean age = 11.64, SD = 0.34).

Material

We made an additional set of 24 images containing two empty gumball machines and groups of green and purple marbles to the side of each machine. We created images with the same numbers of marbles as each of the 24 trials in the ‘exact numerical value probability task’ in Experiment 1. Figure 4 presents an example image along with two example images used to provide feedback for the participant’s choice.

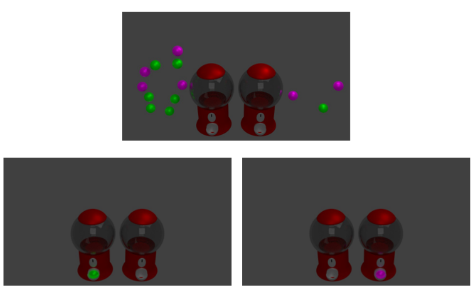

Figure 4: Example images used for presentation and feedback in both the ’conflict’ and ’test’ phases

Procedure

Experiment 2 consisted of three phases. First, children performed the ‘exact-numerical-value- probability task’ in the ***assessment*** phase in order to identify the child’s strategy. The Matlab program recorded the participant’s choices and determined the participant’s strategy score. Children using the ‘greater win’ strategy were coded as ‘1’, ‘lower loss’ strategy was coded as ‘2’, ‘greater difference’ strategy was coded as ‘3’ and the correct proportional strategy was coded as ‘4’.

In the ***conflict phase***, children were semi-randomly assigned to one of two conditions in which they were given feedback about their choices. For each group of strategy users, children were assigned evenly and pseudo-randomly into ‘high conflict’ or ‘low conflict’ conditions consisting of 12 trials. We chose this method to ensure that there were an equal number of children using each strategy in both high and low conflict conditions. Children were told that in this part of the game they will get to see what color marble they get by looking in the tray of the gumball machine that they chose. Since there was no effect of target color in either task of Experiment 1 we decided that all participants would be asked to collect green marbles. Importantly, feedback was given deterministically, meaning that if the child made the mathematically correct choice, they receive a green marble and if they chose incorrectly they received a purple marble. The set of 12 conflict trials were matched to the strategies children used such that if the child used their strategy on every trial they would receive 12 purple marbles and thus children in this condition experience higher conflict between the predictions of their strategy and the actual outcomes. In contrast, children in the low conflict condition as well as children who used the correct proportional strategy during the ***assessment phase*** were given 12 trials randomly selected from the set of 24 trials. Due to the random trial presentation, some trials in the low conflict condition will conflict with their strategy and provide negative feedback while other trials are in agreement with their strategy and provided positive feedback. Importantly, the low conflict condition is an example of an active learning scenario in which the trials are a random assortment of the possible trials. In contrast, the high conflict condition represents a guided learning scenario in which the teacher (in this case the Matlab program) knows the child’s level of understanding and provides the type of examples necessary for the child to overcome their errors.

Finally, during the ***test phase***, the children were asked to play 4 more trials in which they can win prizes depending on how many green marbles they get. Before the beginning of the test phase, children are reminded that they should count the number of marbles in all of the groups and that they can take as long as they need to make a decision. Children’s responses were recorded and all participants received 2 prizes to thank them for participating regardless of the number of green marbles they collected. In this preliminary task we decided to present children with 4 test trials rather than the full set of trials used in the ***assessment phase*** in order to keep the overall time for the experiment below 20 minutes in length.

Experiment 2 Results

Results from the ***assessment phase*** indicated that the majority of children in Experiment 2 utilized one-dimensional strategies. 27 children (57.45%) used the ‘greater win’ strategy, 5 children (10.64%) used the ‘lower loss’ strategy, 8 children (17.02%) used the ‘greater difference’ strategy, and 7 children (14.89%) used the correct proportional strategy. Figure 4 presents the proportion of children using each strategy.

In order to compare children in high and low conflict conditions we calculated the average number of correct responses for each child in both the ***assessment phase*** and ***test phase***. Importantly, children who used the correct proportional strategy were not included in the analyses of the ***conflict*** and ***test phases*** because they could not be assigned to a high conflict condition. For the ***assessment phase***, children in the high conflict condition were not significantly different from those in the low conflict condition ($\Delta M = 0.03$, 95% CI $[-0.13, 0.07]$, $t(35.50) = -0.62$, $p = .537$). However, during the ***test phase***, children in the high conflict condition (74% correct) performed significantly better than children in the low conflict condition (32% correct; $\Delta M = -0.42$, 95% CI $[0.26, 0.58]$, $t(36.63) = 5.29$, $p < .001$). Figure 5 presents the average performance of children in both conflict conditions.

Experiment 2 Discussion

Results from the ***conflict*** and ***test phases*** indicated that children were able to switch strategies after being provided with enough negative feedback using trials which conflicted with their strategy suggesting that younger children are capable of using the correct proportional strategy if they are provided with enough evidence that their original strategy is not working. Previous research has investigated the influence of instruction and feedback on children's understanding of probability. Fischbein et al. (1970) presented 5- to 13-year-old children with a similar 2AFC random draw task. On trials containing the same ratio of marbles, younger children systematically chose the distribution with the larger number of target objects. Following instruction, performance on these trials increased to chance levels. However, since Fischbein et al. (1970) did not assess children's strategies their instruction conditions were not tailored to the child's prior understanding of probability. Furthermore, when performance is at chance level, it is difficult to discern whether children had learned the correct strategy or whether they were choosing randomly.

Note that in Experiment 2 feedback was provided deterministically in both conditions. While we view this as an important control for comparing the high and low conflict conditions we recognize this this method limits the ecological validity of the feedback tasks. Falk et al. (2012) investigated whether children will change their choice after viewing the outcome of a random draw and thus children in their task were provided with probabilistic feedback. Results revealed that children's choices were less consistent following a losing draw compared to a winning draw and that this difference declined with age. However, since children were only presented with each trial twice, the authors did not investigate whether this feedback influenced their overall strategy. Future work will address the influence of probabilistic and deterministic feedback on children's probability judgments.

General Discussion

Findings from Experiments 1 and 2 suggest that children from the US progress slower in their understanding of the proportional nature of probability compared to children in Israel (Falk et al., 2012). Importantly, children in the current US sample appear to use the correct proportional strategy around the same age that the Common Core State Standards suggest they should be formally introduced to probability in school (Best Practices, 2017). In Experiment 2 we demonstrate that feedback influences children's choice strategy but only when they are provided with decisions in which a correct choice conflicts with their strategy's choice.

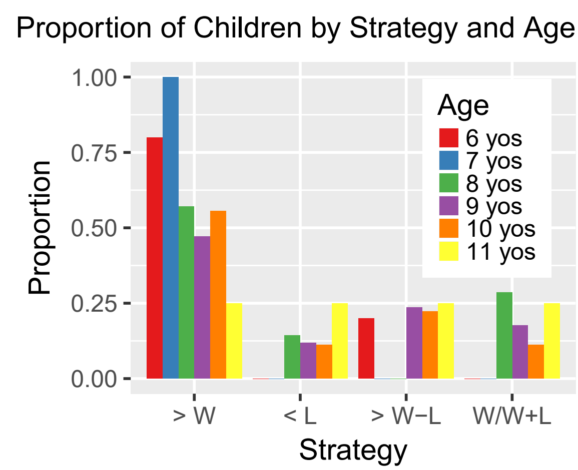

Figure 5: Proportion of children using each strategy by age group.

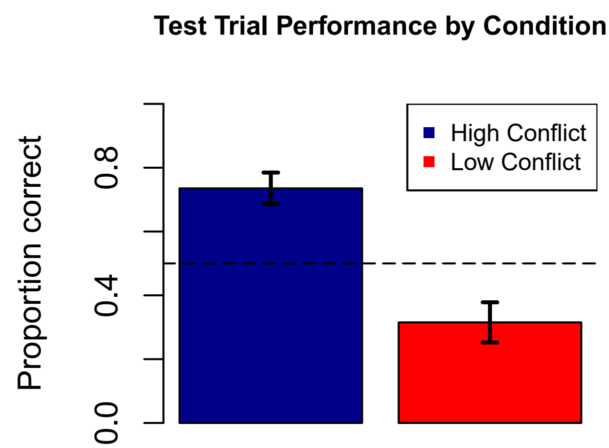

Figure 6: Average test trial performance by condition.

The literature on probabilistic reasoning suggests that even infants have an intuitive understanding of the proportional nature of probability (Denison & Xu, 2014; Teglas et al., 2007; Xu & Garcia, 2008) yet findings from the current study indicate that older children make errors similar to those reported on the 'whole number bias' in fraction learning (Ni & Zhou, 2005). According to the Integrated Theory of Mathematical Development (Siegler, 2016) children learn rational numbers by analogy to whole numbers. From this perspective, the 'whole number bias' occurs when they inappropriately extend whole number properties to rational numbers. Indeed, previous research on proportional reasoning suggests that children can make accurate proportional match choices when provided with continuous proportions compared to discretized proportions (Boyer et al., 2008) and it is believed that familiarity with counting objects plays a role in their errors with discretized proportions. We suggest that a similar 'whole number bias' is at work in explaining the errors children make in our probability tasks. When children are given the right amount of feedback (as in Experiment 2) they realize that their whole number strategy is wrong and are able to fall back on their intuitive sense of probability, using proportional reasoning correctly.

When children enter the classroom, they do not enter as a blank slate, rather they bring with them their intuitions and prior beliefs about a particular domain. Modern constructivist theories draw on concepts from Bayesian probability to express developmental change as the integration of prior beliefs with new information (Fedyk & Xu, 2017; Gopnik & Wellman, 2012). One of the critical duties of a teacher is identifying what a learner already knows in order to design appropriate instruction. Indeed, previous research has found that the degree to which a teacher draws out and expands upon a student's mathematical knowledge can influence a learner's understanding of part-whole relations in fraction representations of mathematical problems (Saxe, Gearhart, & Seltzer, 1999). This 'guided learning' approach is critical in probability learning because the uncertainty of outcomes adds noise to the learning signal. In Experiment 2 we demonstrate that children change their probabilistic decision making strategy when provided with examples that disconfirm their prior beliefs about probability (high conflict condition) but not when given a mixed set of confirming and disconfirming examples (low conflict). These results suggest that a learner's prior beliefs about probability influence how they respond to feedback.

Acknowledgements

We thank the research assistants, lab personnel and graduate students in the Berkeley Early Learning Lab. This research was funded by the National Science Foundation Graduate Research Fellowship under Grant No. (DGE 1106400) S.M.O., and an NSF grant to F. Xu (#1640816).

References

Best Practices, N. G. A. C. for. (2017, december). Common core state standards. Retrieved from `http://www.corestandards.org/`

Boyer, T. W., Levine, S. C., & Huttenlocher, J. (2008). Development of proportional reasoning: Where young children go wrong. *Developmental Psychology*, *44*(5), 1478–1490.

Chapman, R. H. (1975). The development of children's understanding of proportions. *Child Development*, *46*, 141–148.

Denison, S., & Xu, F. (2014). The origins of probabilistic inference in human infants. *Cognition*, *130*(3), 335–347.

Falk, R., Yudilevich-Assouline, P., & Elstein, A. (2012). Children's concept of probability as inferred from their binary choices—revisited. *Educational Studies in Mathematics*, *81*(2), 207–233.

Fedyk, M., & Xu, F. (2017). The epistemology of rational constructivism. *Review of Philosophy and Psychology*, 1–20.

Fischbein, E., Pampu, I., & Mnzat, I. (1970). Comparison of ratios and the chance concept in children. *Child Development*, *41*, 377–389.

Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: Causal models, bayesian learning mechanisms, and the theory theory. *Psychological Bulletin*, *138*(6), 1085.

McCrink, K., & Spelke, E. S. (2016). Non-symbolic division in childhood. *Journal of Experimental Child Psychology*, *142*, 66–82.

McCrink, K., & Wynn, K. (2007). Ratio abstraction by 6-month-old infants. *Psychological Science*, *18*(8), 740–745.

Mix, K. S., Levine, S. C., & Huttenlocher, J. (1999). Early fraction calculation ability. *Developmental Psychology*, *35*(1), 164–174.

Ni, Y., & Zhou, Y.-D. (2005). Teaching and learning fraction and rational numbers: The origins and implications of whole number bias. *Educational Psychologist*, *40*(1), 27–52.

O'Grady, S., & Xu, F. (submitted). The development of non-symbolic probability judgments in children the development of non-symbolic probability judgements in children. *Child Development*.

O'Grady, S., Griffiths, T. L., & Xu, F. (2016). Do simple probability judgments rely on integer approximation? In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.), *Proceedings of the 38th annual conference of the cognitive science society.* Philadelphia, PA: Cognitive Science Society.

Piaget, J., & Inhelder, B. (1975). *The origins of the idea of chance in children*. New York, NY: Norton & Company.

Saxe, G. B., Gearhart, M., & Seltzer, M. (1999). Relations between classroom practices and student learning in the domain of fractions. *Cognition and Instruction*, *17*(1), 1–24.

Siegler, R. S. (2016). Magnitude knowledge: The common core of numerical development. *Developmental Science*, *19*(3), 341–361.

Sophian, C., & Wood, A. (1997). Proportional reasoning in young children: The parts and the whole of it. *Journal of Educational Psychology*, *89*(2), 309–317.

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*, *104*(48), 19156–19159.

Xu, F., & Garcia, V. (2008). Intuitive statistics by 8-month-old infants. *Proceedings of the National Academy of Sciences*, *105*(13), 5012–5015.

Yost, P. A., Siegel, A. E., & Andrews, J. M. (1962). Nonverbal probability judgments by young children. *Child Development*, *33*, 769–780.