Learning the Proportional Nature of Probability from Feedback

Shaun O'Grady
shaun.ogrady@berkeley.edu
Department of Psychology

Geoffrey Saxe
saxe@berkeley.edu
Department of Education
University of California Berkeley

Fei Xu
fei_xu@berkeley.edu
Department of Psychology

Abstract

People make decisions based on probabilistic information every day and often use inaccurate, heuristic decision rules. Although a great deal of research has investigated the developmental trajectory of accurate probability judgments, very little research has investigated how the learning process unfolds. In the current study a microgenetic experimental design was deployed to investigate the influence of feedback on children's probabilistic decision making strategies. Seven- to ten-year-old children (N = 50) first performed a computer-based task to assess the type of strategy they use in a probabilistic judgment task. Next, children receive feedback on a series of 24 trials and then perform a post-test consisting of the same computer-based strategy assessment. Findings revealed that some strategies may benefit from feedback more than others. These results suggest that children can learn about the proportional nature of probability from feedback alone and that the amount and type of feedback influence the learning process.

Keywords: probabilistic reasoning; numerical cognition

Introduction

Every day, children and adults are presented with decisions involving probabilistic information and decades of research has shown that their decisions are often influenced by heuristic biases (Kahneman, 2011; Kahneman & Tversky, 1973; Tversky & Kahneman, 1983). Even on tasks involving simple random draws in which a decision maker is asked to choose between two different options with known probabilities both children and adults have been observed using heuristic decision rules (Falk, Yudilevich-Assouline, & Elstein, 2012; O'Grady & Xu, 2019; Pacini & Epstein, 1999). Although these studies have reported a wealth of data on the factors which promote and inhibit biased decision making, and developmental research has reported a large number of experiments on the developmental trajectory of probabilistic reasoning, very little research has investigated the role of outcome feedback (i.e. the result of a random draw) on children's use of heuristics in probabilistic decision making tasks. In the current paper, a microgenetic experimental design is used to assess the influence of feedback on children's decision making strategies in a simple random draw task.

Developmental trajectory of probabilistic reasoning

Developmental research originally devised by Piaget & Inhelder (1975) studied children's choices in a 2-alternative forced-choice (2AFC) random draw task in which they are presented with two groups of marbles, each contain both red and white marbles in different amounts and are asked to choose the group that is best for getting a target color marble. Piaget and Inhelder (1955) found that young children rely on heuristic decision rules, such as 'pick the group with the greatest number of target outcomes' when making these simple probability judgments and argued that these decision biases suggest children have difficulty with part-whole reasoning, often making comparisons of parts (i.e. making a simple comparison about the number of favorable marbles in each group) without taking into account the relation between the part and the whole (i.e. calculating the proportion of favorable to total outcomes). Although decades of research led to incremental improvements in Piagetian methods in this subfield of developmental psychology (Chapman, 1975; Falk, Falk, & Levin, 1980; Fischbein, Pampu, & Mnzat, 1970), the most thorough and recent experiment conducted by Falk et al. (2012) provides the greatest insight into children's probabilistic reasoning abilities.

Falk et al. (2012) devised a 2AFC random draw task involving a series of 24 trials in which 4- to 11-year-old children were presented with a choice between two groups of marbles each containing a number of favored and unfavored marbles. Children were then asked to choose the group with the best chance of yielding their favored color marble from a single random draw. The 24 trials were designed to discern between four of the most common strategies children have been shown to use in similar tasks. These strategies are (1) 'pick the group with more favorable marbles' ('more favorable') or (2) 'pick the group with the least unfavorable marbles' ('less unfavorable') (3) 'pick the group with the largest difference between favorable and unfavorable marbles' ('greater difference') and (4) the formally correct strategy of 'pick the group with the highest proportion of favorable marbles' ('greater proportion'). Findings from a series of experiments revealed that children progress from 1-dimensional strategies in which they focus on either favorable or unfavorable outcomes (strategies 1 & 2) to more complicated, 2-dimensional strategies in which they attend to both favorable and unfavorable outcomes (strategies 3 & 4). Results revealed that children begin to use the formally correct, proportional strategy around 8 years of age.

Children's difficulties with fraction representations of rational number are notorious and the errors are so common that they are often termed the 'whole number bias' (Ni & Zhou, 2005; Siegler, 2016). This bias presents itself in several different ways across many types of tasks and is very similar to the errors children make in part-whole reasoning during 2AFC random draw tasks. For example, when children are asked to choose the greater of two fractions (say, 1/3 or 2/7) they often choose incorrectly based on comparisons of either the numerator or the denominator (in the above example, choosing 2/7 because the numbers are larger).

Teaching Children Probability Concepts

In a review of the research on statistical edcuation, Garfield & Ahlgren (1988) argue that school-age children have difficulty developing an intuitive understanding of fundamental topics in probabiltiy and statistics for three reasons. First, students have difficulty reasoning about rational number and proportions. Second, probability concepts often conflict with students' real-world experience. Finally, Garfield & Ahlgren (1988) argue that students often develop an aversion to statistics and probability because they learn about these concepts at a very abstract and formal level. Previous research has addressed several of these concerns in their attempts to improve children's understanding of probability.

In formal mathematics, probability is represented as a rational number between 0 and 1, and in the 2AFC random draw task these probabilities are computed as proportions of favorable outcomes. Several groups of researchers have attempted to teach children strategies for calculating and reasoning about probability. Fischbein & Gazit (1984) investigated the effect of teaching probability on 10-13 year olds' predictions of probabilistic outcomes using survey questions about the results of rolling two dice. Students in the experimental group received 12 lessons in which they were taught computational strategies and conceptual relationships in probability. Interestingly, the results revealed that while the experimental group outperformed the control group on computational questions (i.e. questions in which children needed to apply a specific algorithm to identify the correct answer), there was no significant difference in performance on conceptual questions (i.e. questions in which children needed to generalize a concept to a novel context). Although Fischbein & Gazit (1984) report the use of a successful intervention, it is possible that the children in the experimental group merely learned the computational algorithms for solving probability problems without changing their prior concepts about part-whole relations in probability.

Using a didactic approach, Castro (1998) taught young high-school students (14-15 year olds) formal probability. This method encouraged the teachers to incorporate student's intuitive understanding of probability into lessons by allowing students to reflect on their experiences. The experimental group demonstrated significantly more improvement from pre-test to post-test on both probability-reasoning and probability calculation tests. An analysis of the amount of children who changed their answers on similar questions from pre-test to post-test revealed that there were more students who switched their answers in the conceptual change group than in the traditional teaching method group. However, since this study included older teens who may have had experience with formal probability, it is impossible to tell if the conceptual change was a result of the teaching method alone or the interaction of prior conceptual understanding and instruction.

In an intervention study, Nunes, Bryant, Evans, Gottardis, & Terlektsi (2014) attempted to teach 10-year-old children about the importance of understanding the sample space when calculating probability. Participants in the experimental group participated in seven, 50-minute lessons on sample space and probability led by a researcher. Another condition received lessons in mathematical problem solving while a control group of children stayed in the classroom and received regular lessons from their teachers. All three groups received four assessments on understanding sample space, a pre-test, and three post-tests given at various points throughout the program. The experimental group outperformed the two control groups on all three post-tests.

Calculating probability based on proportion of outcomes in the sample space can be accomplished through the use of several cultural forms such as absolute number, ratios, fractions, proportions and percentages. The function of calculating probability is not inherent in any one of these forms and the decision to use one form over another entails a complex interaction of social and individual factors as well as aspects of the problem for which the chosen form is recruited. Nunes and colleagues (2014) argue that ratio representations may be better suited for teaching probability to 9-11 year olds because children understand ratios earlier than proportions and most probability problems are based on proportional judgments. However, since ratio judgments only provide part-part comparisons they may prime children to make correspondences between two different quantities rather than integrating the two quantities by using proportions.

Prior knowledge and instructional context

Discordant assumptions about communicative exchanges can be problematic during instruction if a teacher and a learner view the same forms as supporting different functions (Saxe, 2004). Using a quasi-experimental design, Saxe, Gearhart, & Seltzer (1999) investigated the influence of children's prior understanding of fractions and classroom practices on mathematics learning. Children were categorized as either having or not having a rudimentary part-whole understanding of fractions and classrooms were rated on a scale of alignment with reform standards. High alignment was characterized by the degree to which a teacher draws out and expands upon a student's mathematical knowledge as well as the extent to which conceptual issues are highlighted during problem solving tasks. Importantly, classrooms that espouse either self-discovery or procedural memorization would be considered low in alignment with reform policies. Results revealed that high classroom alignment with reform standards predicted greater performance on a post-test requiring a conceptual understanding of fractions and this effect was stronger for children without a rudimentary understanding of fractions. Interestingly, there was no clear relationship between classroom alignment with reform standards and performance on computational problems regardless of students' prior understanding of fractions. With low levels of alignment to reform principles students without a rudimentary understanding had no basis with which to structure their goals and may have relied on their prior conceptual understanding of integers. However, with supportive classroom environments in which teachers seek to draw-out and build upon a learner's prior knowledge, children can more easily engage with mathematical goals and stand a better chance of learning.

These findings highlight the importance of both prior knowledge and instructional context on a child's ability to learn mathematics. Educators have long understood the importance of providing children with specific feedback based on their prior conceptual knowledge. Indeed, Saxe et al. (1999) found that fraction learning outcomes are a function of a learner's prior understanding (whole number vs rudimentary fraction understanding) and instructional context. Teachers who are able to identify a child's prior conceptual understanding of fractions can construct a learning environment that either confirms accurate conceptualization or scaffolds the learner towards a more thorough conceptualization.

Rationale for the current study

Can children learn to avoid whole number biased choices in probability tasks when they are provided with feedback about the outcomes of their choices? What features of the instructional context allow them to override their non-proportional strategies? We hypothesize that children require consistent feedback on problems which conflict with their prior knowledge and we predict that children who are provided with such feedback will be more likely to reject their incorrect strategy compared to children who are provided with a mix of conflicting and non-conflicting examples.

Methods

Participants

The current experiment was pre-registered (`http://aspredicted.org/blind.php?x=mp6gc9`) with a target sample of 80 children between the ages of 7 and 10 (20 children in each age group: 7-year-olds, 8-year-olds, 9-year-olds, and 10-year-olds), which was determined based on previous research using a similar task (Falk et al., 2012). Currently, data have been collected from N = 50 children (19 7-year-olds, Mean age = 7.5, SD = 0.26; 13 8-year-olds, Mean age = 8.45, SD = 0.23; 9 9-year-olds, Mean age = 9.28, SD = 0.2; and 9 10-year-olds, Mean age = 10.2, SD = 0.2). All fifty children participated in the first session and three children declined to participate in the follow-up session 1 week later (1 7-year-old, 1 8-year-old, and 1 10-year-old).

Material

Images depicting two gumball machines and two groups of green and purple marbles were rendered using Blender (Version 2.78) 3D animation software. Following Falk et al. (2012), each trial image was internally labeled with the trial type designators 'GGGG', 'GGGS', 'SSSG', and 'SSSS' with each letter representing the dimension of comparison and the letter itself relating the correct choice (higher probability of yielding the child's favored color marble) to the incorrect choice (lower probability of yielding the child's favored color). For each target color (i.e. green or purple), two sets of 24 images were created using the same distributions used by Falk et al. (2012) for a total of 96 images.

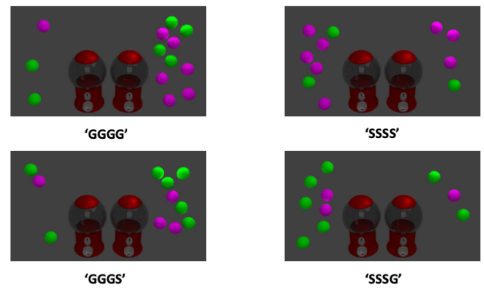

Figure 1: Example images for each of the 4 trial types. In all 4 images, the correct choice for obtaining a purple marble is located on the right side of the image.

Figure 1 presents an example image for each trial type. Note that the correct choice in the figure on the top left (labeled 'GGGG') has a greater amount of favored marbles (1st G), a greater amount of non-favored marbles (2nd G), a greater total of favored and non-favored marbles (3rd G) and a greater difference between favored and non-favored marbles (4th G). In contrast, the correct choice for the image on the top right (labeled 'SSSS') has a smaller amount of marbles in each of these categories compared to the incorrect choice. Children using a strict 'more favorable' strategy would make a correct choice on all 12 'GGGG' and 'GGGS' trials but would choose incorrectly on all 12 'SSSS' and 'SSSG' trials. A child using a strict 'less unfavorable' strategy would make a correct choice on all 12 'SSSS' and 'SSSG' trials but would choose incorrectly on all 12 'GGGG' and 'GGGS' trials. Children using a strict, 'greater difference' would make a correct choice on all 12 'GGGG' and 'SSSG' trials but would choose incorrectly on all 12 'SSSS' and 'GGGS' trials. Finally, a child using the formally correct proportional strategy would choose correctly on all 24 trials.

Procedure

Children were seated approximately 60 cm away from a MacBook Pro laptop (OSX; Screen resolution 1280 x 800) and told they would play a game in which they would try to collect green or purple marbles from one of two different gumball machines. The task consisted of a self-paced game automated using the psychophysics toolbox written for the MatLab programming language (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997). In order to maintain an average testing time of 20 minutes, the experiment was split into 2 testing sessions spaced 1 week apart. Children completed the *assessment phase* during session 1 and then completed the *conflict phase* and *post-test phase* during session 2.

Assessment Phase During the first testing session, the experimenter explained the task and children were prompted to choose their favorite of the two colors, either purple or green.

For each of the 24 images, the computer presented the image at random along with 4 counting prompts, one for each group of marbles (i.e. "How many (green/purple) marbles are on this side (left/right)?"). The child responded by pressing the appropriate number key on the keyboard. An error message was presented if the child chose the wrong number and the game did not progress until the child pressed the correct number key. Counting prompts for each color and side were randomized for each image. Once children completed the counting prompts, they were prompted with the question "Which would you pick to get a (green/purple) marble?". Importantly, the position of the marbles on the screen were randomized to prevent children from choosing based on the positions of their favorite color marble. However, in order to ensure that children did not rely on the placement of the marbles on the screen they were told to "Pretend that the marbles will go into the machines and that the machines will be shaken up so you don't know what's going to come out next."

Following the methods outlined by Falk et al. (2012), the MatLab program discerned which strategy the child used based on their performance on each of the four trial types. After children completed the *assessment phase*, the MatLab program calculated point scores for each strategy based on the choices that the child made. Whichever strategy had the highest point score was deemed to be that child's strategy. Point scores could range from 0 to 24 with 0 indicating no strategy-consistent responses and 24 indicating perfect strategy use. Participants using the 'more favorable' strategy provided about 21 (M = 21.6; SD = 3.5) out of 24 strategy-consistent responses, while those using the 'less unfavorable' strategy provided 16 (M = 16; SD = 1.79) out of 24 strategy-consistent responses, participants using the 'greater difference' strategy provided 20 (M = 20.2; SD = 20.2) out of 24 strategy-consistent responses, and participants using the 'greater proportion' strategy provided 19 (M = 19.78; SD = 19.78) out of 24 strategy-consistent responses.

Conflict Phase Children were semi-randomly assigned to one of two different conditions ensuring that an equal number of children using each strategy were assigned to both conditions. In the 'half-conflict' condition, children viewed all 24 trials, 12 of which conflicted with the child's strategy and 12 of which did not conflict. Children in the 'high-conflict' condition viewed 24 trials that conflicted with their strategy.

In the 'high conflict' condition, feedback trials were assigned as follows. Children designated as using the 'more favorable' strategy viewed 12 'SSSS' trials and 12 'SSSG' trials. Children using the 'less favorable' strategy viewed 12 'GGGG' trials and 12 'GGGS' trials. A child using the 'greater difference' strategy viewed 12 'SSSS' trials and 12 'GGGS' trials while children using the proportional strategy were simply assigned to the 'half-conflict' condition as none of the trials conflicted with their strategy. In all conditions and for all trials, children received feedback in the form of either a favored or unfavored color marble returned in the dispenser of the machine they chose. Importantly, all feedback was provided deterministically, meaning that if a child chose strictly according to their non-proportional strategy in the 'high-conflict' condition, they would receive 24 unfavored marbles and a child in the 'half-conflict' condition would receive 12 favored and 12 unfavored marbles. Children in the 'half-conflict' condition received a mix of confirmatory and dis-confirmatory feedback with respect to their strategy while children in the 'high-conflict' condition received only dis-confirmatory feedback.

Post-test phase After completing the *conflict phase*, each child received the same 24 trials they viewed in the *assessment phase* one week prior in a randomized order. Importantly, the *post-test phase* is an immediate post-test because it occurred directly following the *conflict phase*.

Results

The results of each phase of the experiment are reported separately below along with a brief discussion section. For all three phases, analyses consisted of comparisons of Generalized Linear Regression Models with Mixed effects (GLMMs) using the lme4 package written for the R statistical programming language (Bates, Maechler, Bolker, & Walker, 2015). All models predicted the binary response variable while holding participant ID as a random effect. Nested models were compared using Chi Squared tests for model fits while non-nested models were compared using the Akaike Information Criterion (AIC), a measure of model fit in which models with smaller AICs are preferred over models with higher AICs. For all three phases of the experiment, modeling results revealed no influence of participant gender, favored color, on performance. Model coefficients for GLMMs are reported as log-odds, that is, the log of the odds ratio of correct to incorrect responses.

Assessment Phase Results

In order to investigate the influence of age on strategy use we used a Chi Squared test to assess the independence of age group and strategy. Results of the χ^2 test revealed that older children were significantly more likely to use the correct proportional strategy ($\chi^2(9, n = 50) = 18.11$, $p = .034$). Figure 2 presents the proportions of children using each strategy by age. Note that the two younger age groups (7-year-olds and 8-year-olds) are predominantly relying on the, 'more favorable' strategy whereas children in the two older age groups (9-year-olds and 10-year-olds) have a more equal spread across the four different strategies.

Comparisons of GLMMs revealed that the model with the best fit to the assessment phase data was the model predicting performance from strategy alone ($AIC_{Strategy}$ =1562.96). This model outperformed the null model (AIC_{null} = 1613.55; χ^2= 56.6; $df = 3$; $p < .001$), as well as the model predicting performance from age (AIC_{Age} =1609.59). More complex models predicting performance from age and strategy ($AIC_{Strat+Age}$ = 1561.79; χ^2= 3.17; df = 1; p = .07) and the interaction of age and strategy ($AIC_{Strat*Age}$ = 1564.95; χ^2= 6.01; df =

4; p = .20) did not perform better than the model for strategy alone. Thus, the simpler model is preferred since it can predict the same amount of variance with fewer model parameters. There was no effect of trial number indicating that children's performance did not improve with time during the *assessment phase*.

Inspection of model coefficients reveals that the log-odds of a correct reponse increased for children using the 'greater difference' ($\beta_{>F-U}$ = 0.53; SE = 0.2; 95% CI [0.13, 0.93]), and 'less unfavorable' strategies ($\beta_{<U}$ = 0.17; SE = 0.18; 95% CI [-0.19, 0.53]), as well as those using formally correct proportional strategy ($\beta_{>F/F+U}$ = 1.49; SE = 0.19; 95% CI [1.11, 1.87]) compared to children using the 'more favorable' ($\beta_{>F(Intercept)}$ = 4.003^{-16}; SE = 0.07; 95% CI [-0.09, 0.2]). However, only the coefficients for 'greater difference' and 'greater proportion' strategies reached statistical significance (Wald test: 'greater difference': $p < .01$; 'greater proportion': $p < .001$).

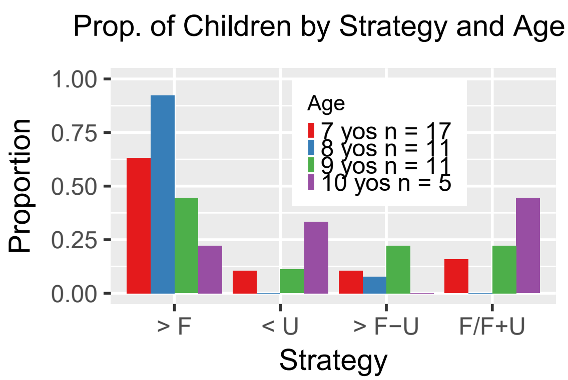

Figure 2: Proportion of children using each strategy by age group. Strategies are designated as follows: '> F': more favorable; < U: less unfavorable ; > F-U: greater difference; > F/F+U: greater proportion.

Assessment Phase Discussion

Results of the current study converge with those of previous reports indicating that children's use of the correct proportional strategy improves with age (Falk et al., 2012; O'Grady & Xu, 2018). Importantly, results of the GLMM comparisons revealed an effect of strategy on performance indicating that children who attended to the number both of favorable and unfavorable marbles in each choice performed better than children who made their choices based on one single dimension (i.e. choosing based solely on the number of favorable or unfavorable marbles).

Conflict Phase Results

Nine children were found to be using the formally correct proportional strategy during the assessment phase the previous week. Since there are no trials that conflict with this strategy, data from these children were excluded from the conflict phase analyses resulting in a sample size of N = 38. Comparisons of GLMMs for this subsample revealed that the model with the best fit to the data predicted performance from the conflict condition and the trial number as well as the interaction between the two variables ($AIC_{Condition*Trial}$ =1143.09). This model outperformed the null model (AIC_{null} = 1613.55; χ^2= 29.74; df = 3; $p < .001$) as well as the simpler models predicting performance from conflict condition ($AIC_{Condition}$ = 1158.39; χ^2= 19.3; df = 2; $p < .001$) and trial number alone (AIC_{Trial} = 1162.96; χ^2= 23.87; df = 2; $p < .001$) and the more complex model accounting for both the conflict condition and trial number without an interaction ($AIC_{Condition+Trial}$ = 1154.51; χ^2= 13.43; df = 1; $p <$ .001). There were no significant effects of the three non-proportional strategies, nor was there an interaction between conflict condition.

Inspection of the model coefficients revealed that the log-odds of a correct decreased for children in the 'high-conflict' condition ($\beta_{High-Conflict}$ = -0.13; SE = 0.35; 95% CI [-0.81, 0.55]) compared to children in the 'half-conflict' condition ($\beta_{Half-Conflict}$ = 0.32; SE = 0.25; 95% CI [-0.18, 0.81]), which is not surprising considering that all of the trials in the 'high-conflict' condition conflicted with the children's strategies whereas only 12 of the 24 trials in the 'half-conflict' condition conflicted with the children's strategies. While the model coefficient for trial number was slightly negative (β_{Trial} = -0.01; SE = 0.01; 95% CI [-0.04, 0.02]) indicating a decrease in the log-odds of a correct response, the interaction between trial number and condition revealed that in the 'high-conflict' condition, trial number had a positive effect on the log-odds ($\beta_{Trial*High-Conflict}$ = 0.08; SE = 0.02; 95% CI [0.04, 0.12]). The interaction between trial order and the 'high-conflict' condition was the only model coefficient to reach statistical significance (Wald test: $p < .001$) indicating that performance improved over time in the 'high-conflict' condition suggesting that children in this condition may have learned from feedback on earlier trials.

Conflict Phase Discussion

Results revealed that both conflict condition and trial order had an effect on performance. Importantly, the interaction between conflict condition and trial number produced the greatest positive effect on performance while the coefficient for the 'high-conflict' condition alone had a negative effect on performance. This set of results suggests that children in the 'high-conflict' condition began by choosing according to their strategy but then switched to another strategy after several trials in which they received negative feedback.

Post-Test Phase Results

Of the 9 children who used the correct proportional strategy during the *assessment phase* only one child (a 10-year-old) did not continue to use the correct proportional strategy. Interestingly, this child used the 'more favorable' strategy and reported that they switched to a simpler strategy because "the game was boring and I wanted to finish it faster" suggesting that this child understood the time-accuracy tradeoff among the various potential strategies. Table 1 presents the number of children using each of strategy in the *post-test phase* ('Post-test' column) based on the child's *assessment phase* strategy ('Assessment' column) and condition ('half-conflict' and 'high-conflict' columns).

Assessment	Post-test	Half-Conflict	High-Conflict
>F	>F	9	3
>F	<U	1	7
>F	>F-U	1	1
>F	>F/F+U	2	4
<U	>F	2	1
<U	<U	0	1
<U	>F-U	0	0
<U	>F/F+U	0	1
>F-U	>F	0	0
>F-U	<U	0	0
>F-U	>F-U	2	0
>F-U	>F/F+U	0	3

Table 1: This table presents the number of children using each strategy listed in the Post-test column during the post-test phase after using the strategy in the Assessment column during the assessment phase.

Comparisons of GLMMs revealed that the model with the best fit to the data predicted *post-test phase* performance based on the interaction between conflict condition and *assessment phase* strategy ($AIC_{Condition*Strategy}$ =1182.6). This model outperformed the null model (AIC_{null} = 1613.55; χ^2= 19.41; df = 5; $p < .001$), the models for conflict condition alone ($AIC_{Condition}$ = 1189.82; χ^2= 15.22; df = 4; $p < .001$) and *assessment phase* strategy alone ($AIC_{Strategy}$ = 1188.48; χ^2= 11.88; df = 3; p = .01) as well as the model for conflict condition and *assessment phase* without any interactions ($AIC_{Condition+Strategy}$ = 1185.81; χ^2= 7.21; df = 2; p = .03).

Inspection of model coefficients revealed that the only model coefficient to reach statistical significance was the interaction between 'greater difference' strategy and the 'high-conflict' condition which increased the log-odds of a correct reponse ($\beta_{'>F-U'*High-Conflict}$ = 1.74; SE = 0.65; 95% CI [0.46, 3.01]; Wald test: $p < 0.01$) compared to the children using the 'more favorable' strategy in the 'half-conflict' condition ($\beta_{Intercept}$ = 0.05; SE = 0.35; 95% CI [-0.64, 0.73]). All remaining coefficients did not reach statistical significance. Figure 3 presents the proportion of correct responses by conflict condition and *assessment phase* strategy.

Post-Test Phase Discussion

The interaction between conflict condition and *assessment phase* strategy indicates that children using the 'greater difference' strategy benefited more from the 'high-conflict' condition compared to children using the other 2 strategies. Although these findings are promising, there were only three children using the 'greater difference' and three children using the 'less unfavorable' assigned to the 'high-conflict' condition thus more data will be needed to make any firm conclusions. However, it is interesting to view the differences between the 'half-conflict' and 'high-conflict' condition for children using the 'more favorable' strategy. Note from table 1 that only 3 of the 15 children using this strategy in the 'high-conflict' condition (20%) continued using their strategy after the feedback condition while 9 of the 13 assigned to the 'half-conflict' condition (69.2%) continued to use the strategy.

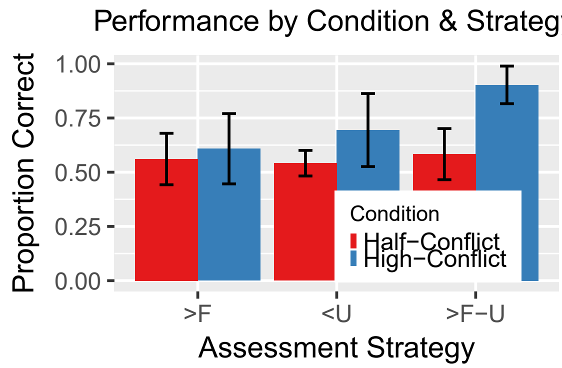

Figure 3: Proportion of correct responses in the post-test phase by conflict condition and assessment phase strategy. Error bars indicate standard deviation.

General Discussion

The current findings provide three important insights into the development of proportional reasoning in probability judgments. First, during the *assessment phase*, children using the correct proportional strategy performed better than children using non-proportional strategies and there was no effect of trial order indicating that children relied on the same strategy throughout the task. These findings provide an important replication of previous research (Falk et al., 2012; O'Grady & Xu, 2018). Second, during the *conflict phase*, results revealed an interaction between trial order and conflict condition indicating that children in the 'high-conflict' condition performed better on later trials compared to earlier trials while this effect was not found in the 'half-conflict' condition. Previous research investigating the influence of feedback in similar decision making tasks have observed the effect of feedback on a single trial (Falk et al., 2012) or presented feedback on a limited number of trials (O'Grady & Xu, 2018). In contrast, the current approach allows for the observation of a comprehensive set of trials allowing for a more thorough understanding of how feedback influences strategy change. Finally, results from the *post-test phase* revealed that while children in the 'high-conflict' condition were more likely to abandon their strategy, children using the 'greater difference' strategy seemed to have gained the most from this feedback. Although more data need to be collected, these preliminary results suggest that strategy-specific feedback can help children overcome 'whole number bias' in probability tasks.

Our findings shed new light on how children learn about probability. In the 'half-conflict' condition, we attempted to mimic the experience a child would gain from actively exploring the environment. In contrast, the 'high-conflict' condition was meant to provide a learning context tailored to the child's prior understanding of proportional relations in probability.

Constructivist theories of cognitive development highlight the interaction between a learner's prior knowledge and new information gained through their own active exploration as well as through socio-cultural processes like education (Piaget & Inhelder, 1975; Vygotsky, 1962). By assessing the child's prior understanding and then presenting examples which conflict with that understanding, the computer program in the 'high-conlfict' condition is acting much like a constructivist teacher, identifying the learner's prior knowledge, providing them with conflicting evidence, and allowing the child to construct a new conceptual understanding.

Why do children benefit from feedback in the 'high-conflict' condition but not from the feedback on the 12 conflicting trials in the 'half-conflict' condition? Children understand the uncertain nature of probability, that is, they understand that they may receieve their non-favored marble even though they chose the 'best' option for getting their favorite color. In the 'half-conflict' condition this prior understanding of uncertainty allows children to continue to believe their inaccurate strategy is correct because the negative feedback can be chalked up to chance. However, children in the 'high-conflict' condition are forced to reconcile their inaccurate strategy with the evidence at hand. Thus they must reject their prior conceptualization of probability and construct a more accurate representation. Although the current evidence suggests that children in the 'high-conflict' condition reject their prior conceptualization of probability, future research will be necessary to uncover the new representations that children construct as well as the process through which this transition ocurrs.

Acknowledgements

We would like to thank the families who participated in this study as well Berkeley Early Learning Lab staff for their support. 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

Bates, Maechler, Bolker, & Walker. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Brainard. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436.

Castro, C. S. (1998). Teaching probability for conceptual change la enseanza de la probabilidad por cambio conceptual. *Educational Studies in Mathematics*, *35*(3), 233–254.

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

Falk, R., Falk, R., & Levin, I. (1980). A potential for learning probability in young children. *Educational Studies in Mathematics*, *11*(2), 181–204.

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.

Fischbein, E., & Gazit, A. (1984). Does the teaching of probability improve probabilistic intuitions? *Educational Studies in Mathematics*, *15*(1), 1–24.

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

Garfield, J., & Ahlgren, A. (1988). Difficulties in learning basic concepts in probability and statistics: Implications for research. *Journal for Research in Mathematics Education*, 44–63.

Kahneman, D. (2011). *Thinking fast & slow*. New York, NY: Farrar, Strauss,; Giroux.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, *80*, 237–251.

Kleiner, Brainard, & Pelli. (2007). What's new in psychtoolbox-3? *Perception*, *36*.

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.

Nunes, T., Bryant, P., Evans, D., Gottardis, L., & Terlektsi, M.-E. (2014). The cognitive demands of understanding the sample space. *ZDM*, *46*(3), 437–448.

O'Grady, S., & Xu, F. (2018). Whole number bias in children's probability judgments. In *Proceedings of the 40th annual conference of the cognitive science society*.

O'Grady, S., & Xu, F. (2019). The development of non-symbolic probability judgments in children the development of non-symbolic probability judgements in children. *Child Development*.

Pacini, R., & Epstein, S. (1999). The interaction of three facets of concrete thinking in a game of chance. *Thinking and Reasoning*, *5*(4), 303–325.

Pelli. (1997). The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.

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

Saxe, G. B. (2004). Practices of quantification from a sociocultural perspective. *Cognitive Developmental Change: Theories, Models and Measurement*, 241–263.

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.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, *90*(4), 293–315.

Vygotsky. (1962). *Thought and language*. Cambridge, MA: The MIT press.