CHILD DEVELOPMENT

Child Development, January/February 2018, Volume 89, Number 1, Pages 205–218

Probability Learning: Changes in Behavior Across Time and Development

Rista C. Plate, Jacqueline M. Fulvio, Kristin Shutts, C. Shawn Green, and Seth D. Pollak
University of Wisconsin–Madison

Individuals track probabilities, such as associations between events in their environments, but less is known about the degree to which experience—within a learning session and over development—influences people's use of incoming probabilistic information to guide behavior in real time. In two experiments, children (4–11 years) and adults searched for rewards hidden in locations with predetermined probabilities. In Experiment 1, children (n = 42) and adults (n = 32) changed strategies to maximize reward receipt over time. However, adults demonstrated greater strategy change efficiency. Making the predetermined probabilities more difficult to learn (Experiment 2) delayed effective strategy change for children (n = 39) and adults (n = 33). Taken together, these data characterize how children and adults alike react flexibly and change behavior according to incoming information.

Individuals learn to understand associations in their environments by attending to probabilistic information (Gopnik & Wellman, 2012). Extant research indicates that both children and adults track probabilistic information across many situations (e.g., Denison, Bonawitz, Gopnik, & Griffiths, 2013; Saffran, Aslin, & Newport, 1996; Xu & Garcia, 2008). For example, a child might notice that certain actions are highly associated with making friends (e.g., sharing), whereas other actions are less likely to foster friendships (e.g., being bossy). Less is known, however, about the processes whereby individuals use their experience with probabilistic information to organize behavior in real time. We explored the influence of experience on probability learning within two time frames: (a) over the course of an experimental session and (b) across a wide age range (in order to assess accumulated experience over development).

This research was supported by the National Institute of Mental Health through Grant R01MH61285 to Seth D. Pollak and by the National Institute of Child Health and Human Development through Grant R01HD07089053 to Kristin Shutts. Infrastructure support was provided by the Waisman Center at the University of Wisconsin–Madison through the National Institute of Child Health and Human Development (P30HD03352). Rista C. Plate was supported by a National Science Foundation Graduate Research Fellowship (DGE-1256259) and an Emotion Research Training Grant (T32MH018931-24) from the National Institute of Mental Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Mental Health or the National Institutes of Health.

Correspondence concerning this article should be addressed to Rista C. Plate, Department of Psychology, University of Wisconsin–Madison, 1202 West Johnson Street, Madison, WI 53706. Electronic mail may be sent to rplate@wisc.edu.

Probability Learning in Children

Attention to probabilistic information is evident beginning in infancy (e.g., Duffy, Huttenlocher, & Crawford, 2006; Saffran et al., 1996). For example, when 7- to 8-month-old infants watched an experimenter pull a series of balls from a box that they saw contained many more red than white balls (e.g., 70 red and 5 white), they looked longer when the experimenter drew a series of white balls compared to a series of red balls (Xu & Garcia, 2008). Such data presumably reflect infants' understanding that the results were not representative of the colors distributed throughout the box. Preschool-age children also attend to the probability distribution of a sample when making inferences. In one instance (Denison et al., 2013), preschoolers saw that one block from a collection of colored blocks (e.g., 20 red, 5 blue) made a toy light up and play music. Children's guesses regarding which color block activated the toy reflected the distribution of red and blue blocks.

In addition to awareness and tracking of probabilities, young children can also use such information to guide their own behavior. For example, 24-month-olds who observed that one object was highly likely to produce lights and music while another object was less likely to produce lights and music, selected the former object over the latter when given a choice (Waismeyer, Meltzoff, & Gopnik, 2015). In another study (Kushnir, Xu, & Wellman, 2010), preschoolers

© 2017 The Authors
Child Development © 2017 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2018/8901-0020
DOI: 10.1111/cdev.12718

used probability information to respond appropriately to an agent's request for a toy: Those who had previously seen the agent select the least frequent kind of toy from a box later provided that kind of toy to the agent; those who had previously seen the agent select the most frequent kind of toy from a box showed no preference when later selecting a toy for the agent. These demonstrations give us clues about how children come to learn about and act on their environments based on probabilistic information.

As these examples illustrate, research on children's attention to and use of probabilistic information has provided insights into the robustness of these abilities in young children. However, much of the extant research on children's probability learning, including the studies described above, has been limited to tasks in which participants acquire all relevant information prior to acting themselves (exceptions described in subsequent section). For instance, in the aforementioned study by Waismeyer et al. (2015), participants saw the high- and low-probability objects produce outcomes before they were presented with a choice of objects to manipulate themselves. Yet in children's daily lives, they must often distill incomplete or partial information from a fluid environment—initially with few exemplars—and use that information to guide their decisions and actions over time. Additionally, many methods used to assess probability learning in children are designed for young children. Less research has been done using methods that compare across age groups, and researchers have traditionally focused on a narrow range of ages. Therefore, while we see evidence of probability learning in young children, it is not clear whether the approaches young children use in this learning process are similar to those used by more mature individuals, nor is it clear whether children's approaches are stable or change across development.

Measurement of Probability Learning

To better understand children's use of accumulating experience over time and development, we created a probability learning task. Probability learning tasks feature probabilistic information that is revealed to the participant over the course of many trials. In the simplest version of a probability learning task, the participant is repeatedly asked to make a choice between two options with one of the two options being rewarded more frequently than the other (e.g., Option 1 is rewarded on 65% of trials and Option 2 is rewarded on 35% of trials). The participant is not made explicitly aware of these underlying probabilities and thus must use the observed outcomes in order to make effective choices on subsequent trials. Data from these tasks are typically considered as the aggregated behavior of an individual over the entirety of the experiment. Thus, previous research has not been able to inform our understanding of whether individuals change their use of incoming information to improve behavior as they gain experience.

Using data averaged across trials, researchers have identified two types of strategies individuals use in these probability learning tasks: "probability matching" and "maximizing" (Vulkan, 2000). Under a probability matching strategy, participants select options in proportion to the probability that those options will be rewarded. Thus, in the example highlighted in the previous paragraph, a participant exhibiting matching would select Option 1 on approximately 65% of trials and Option 2 on 35% of trials. In contrast, a maximizing strategy would result in participants primarily selecting the option with the higher probability of reward. In the same example, a participant exhibiting maximizing would nearly always select Option 1.

Across myriad variations in task structure and even across species, participants most frequently exhibit matching behavior in probability learning tasks (Vulkan, 2000), with some exceptions noted below. Such behavior is paradoxical because matching results in less reward receipt than maximizing. This is because participants cannot know when a given location or response option will be rewarded, even if they are cognizant of the overall reward rate. In the example above, if a participant completed 100 trials, percent accuracy using a matching strategy would be approximately 55%: 0.65×0.65 (percent of time Option 1 is chosen $\times$ percent of time Option 1 is correct) $+ 0.35 \times 0.35$ (percent of time Option 2 is chosen $\times$ percent of time Option 2 is correct). However, a participant's percent accuracy using a maximizing strategy would be 65%: 1.0×0.65 (percent of time Option 1 is chosen $\times$ percent of time Option 1 is correct).

Developmental Differences in Probability Learning

In contrast to the research described earlier with infants and young children, probability learning tasks allow for direct comparison across ages. However, the probability learning task literature contains conflicting reports of developmental differences, in particular with regard to matching versus maximizing. A number of studies have found that school-age children demonstrate rates of matching similar to

adults and that younger children (ages 3–5 years) demonstrate maximizing more so than older children (e.g., Brackbill & Bravos, 1962; Derks & Paclisanu, 1967). Yet other researchers have found that adults maximize rewards more effectively than children (Moran & McCullers, 1979). Perhaps the developmental differences (or lack thereof) are not clear because these past studies have averaged behavior across the experimental session. Therefore, while research employing standard probability learning tasks differs from those reported earlier (in which children acquire all relevant probabilistic information before acting), they still do not measure continuous behavior change. Rather than characterizing developmental differences in terms of overall average behavior, our experiment aims to understand how individuals use incoming information to direct behavior over time. In doing so, we address two significant gaps in the literature: (a) assessing continuous behavior change and (b) elucidating developmental differences.

Present Research

Using an analytic approach that allows for continuous measurement of behavior on a probability learning task, we can begin to understand whether accumulating more information about underlying probabilistic structure in an environment influences behavior over time. Additionally, by including a wide age range of children and adults, we can assess the role of accumulated experience over development and examine whether probability learning abilities change across childhood and into adulthood. As individuals gain more experience in the world—and encounter situations in which they must monitor associations between actions and events—are they better able to use probabilistic information to direct behavior change in real time?

Participants completed a task where they were presented with multiple options that were rewarded with different probabilities. We recorded participants' choices throughout the experimental session and then assessed participants' patterns of choices. Because previous studies have traditionally averaged participant behavior across trial blocks, it was unclear whether individuals would change behavior within a single experimental session and if so, what patterns of behavior would emerge. However, based on theories of probability learning, if behavior change was captured within the experimental session, we expected early behavior to be consistent with probability matching, but later behavior to be progressively consistent with maximizing. This prediction was based on the premise that matching reflects expectations individuals have at the start of an experiment (i.e., the expectation that outcomes are not generated randomly; Green, Benson, Kersten, & Schrater, 2010) and that matching allows participants to sample multiple options (Denison et al., 2013). However, a matching strategy garners fewer rewards than a maximizing strategy. Consequently, we expected that individuals would turn to maximizing once they had explored other options. Additionally, we suspect that the paradoxical findings in previous research that individuals match instead of maximize reflect that the amount of matching individuals demonstrate early in the experiment outweighs maximizing behavior later in the experiment, thus influencing the average across trials. Again, we aim to present a fine-grained evaluation of behavior in order to shed light on this possibility.

While previous research has demonstrated efficacy of probability learning in young children, it is unclear how this might translate across childhood and compare to probability learning in adulthood. If more mature individuals have gained efficiency through more experience with various types of probabilistic learning environments, we might expect them to demonstrate greater proficiency than young children. Yet, it is unclear whether there should be developmental differences in *ability* to change behavior over time, *pattern* of change over time, and/or *timing* to change behavior over time. Disentangling not only whether there are developmental differences, but also the nature of those differences, will inform our understanding of how individuals of various ages integrate information in their environments over time.

Experiment 1

In Experiment 1, children and adults searched for a reward that was hidden behind one of multiple locations depicted on a computer screen. The probability of obtaining a reward differed by location, as described below.

Method

Participants

The sample included 42 children and 32 young adults (recruited May 2013–March 2015). We tested a wide age range of children to examine potential

differences, not only between children and adults but also across childhood. Children were recruited from a registry of families who had previously expressed interest in participating in research and were from the community in a large Midwestern city. Adults were undergraduates who participated for course credit (at a large Midwestern university) or members of the community recruited via word of mouth from the same community. The Institutional Review Board approved the research. Adult participants and parents of child participants gave informed consent. Children who were 11 years of age gave written assent, and younger children gave verbal assent. Parents received $20 for their time and children chose a prize for their participation. Adults who were not participating for course credit received $20.

Procedure

The experimenter told participants they would play a computer game in which an elf hid coins behind rocks. She then explained that the goal was to find as many coins as possible during the game, and that any one of eight rock locations could be chosen on each trial. On two practice trials, participants either (a) found the coin and received points or (b) failed to find the coin and received no points. In order to motivate the children to make effective choices, the children were told that if they found enough coins, they would get to choose from prizes on a shelf they saw as they entered the laboratory. Regardless of whether the participant's choice was correct, the actual location of the coin was revealed on each trial.

Following the practice phase, participants began the test trials. There were two blocks of 100 trials, separated by a break. On each test trial, eight rocks were displayed with equal spacing along a horizontal line on the computer screen (Figure 1, top). When participants selected the correct location on a trial, a coin appeared in place of the rock they selected (Figure 1, bottom left). When participants selected an incorrect location on a trial, a red "x" appeared in the chosen location and the coin was revealed in the correct location (Figure 1, bottom right). From left to right, the following probabilities defined the likelihood of a coin appearing at each rock location on any given trial: 0%–0%–5%–10%–70%–10%–5%–0% (Figure 2; calculation of reward receipt for matching and maximizing is provided in the Supporting Information). To ensure all participants' experiences were statistically equivalent, the outcomes were predetermined to ensure a match to the location probabilities across the trial blocks (i.e., in each 100 trial block, Rock 5 would be rewarded on exactly 70 trials, Rocks 4 and 6 would be rewarded on exactly 10 trials, etc.). These probabilities were not made directly known to the participants; the probabilities had to be learned via experience with the task.

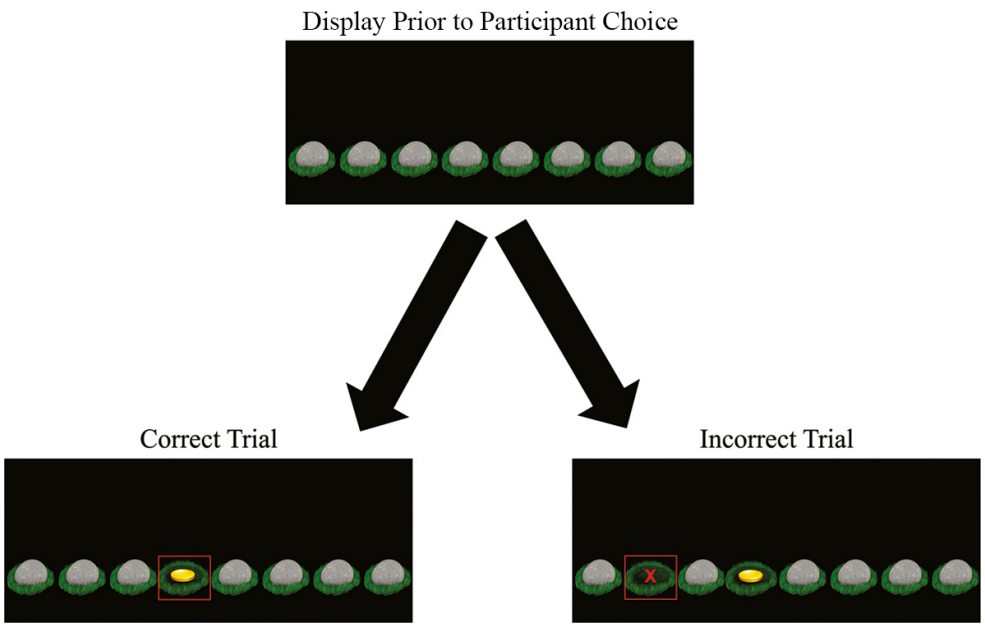

Figure 1. Schematic of the experimental display. Display prior to participant choice (top), and following a correct (bottom left) or incorrect (bottom right) participant choice. [Color figure can be viewed at wileyonlinelibrary.com]

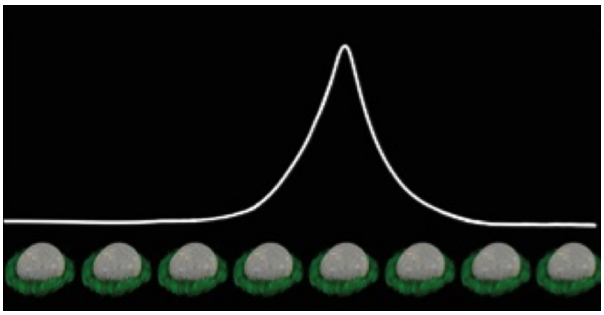

Figure 2. Distribution of rewards by location for Experiment 1. Participants were not shown the explicit probabilities. [Color figure can be viewed at wileyonlinelibrary.com]

Scoring and Analyses

We assessed the extent to which individual participant choices were best captured by one of four different possible models of choice behavior. In brief, the first was a *random choice model*, in which there was an equal and constant probability of the participant selecting each of the eight options. This model assumed participants sampled without a consistent strategy and served as a baseline against which evidence-based models could be compared. The second model was a *probability matching model*. Here participants were expected to choose each option in proportion to the probability that each location had been observed to be correct up to the current trial (Figure 3A). The third model was a *maximizing model* (Figure 3B). Under this model, participants were expected to choose the option that had been observed to have highest probability of reward up to the current trial. The final model was a time-evolving *combination model*, which combined the previous two models. Parameters and fitting

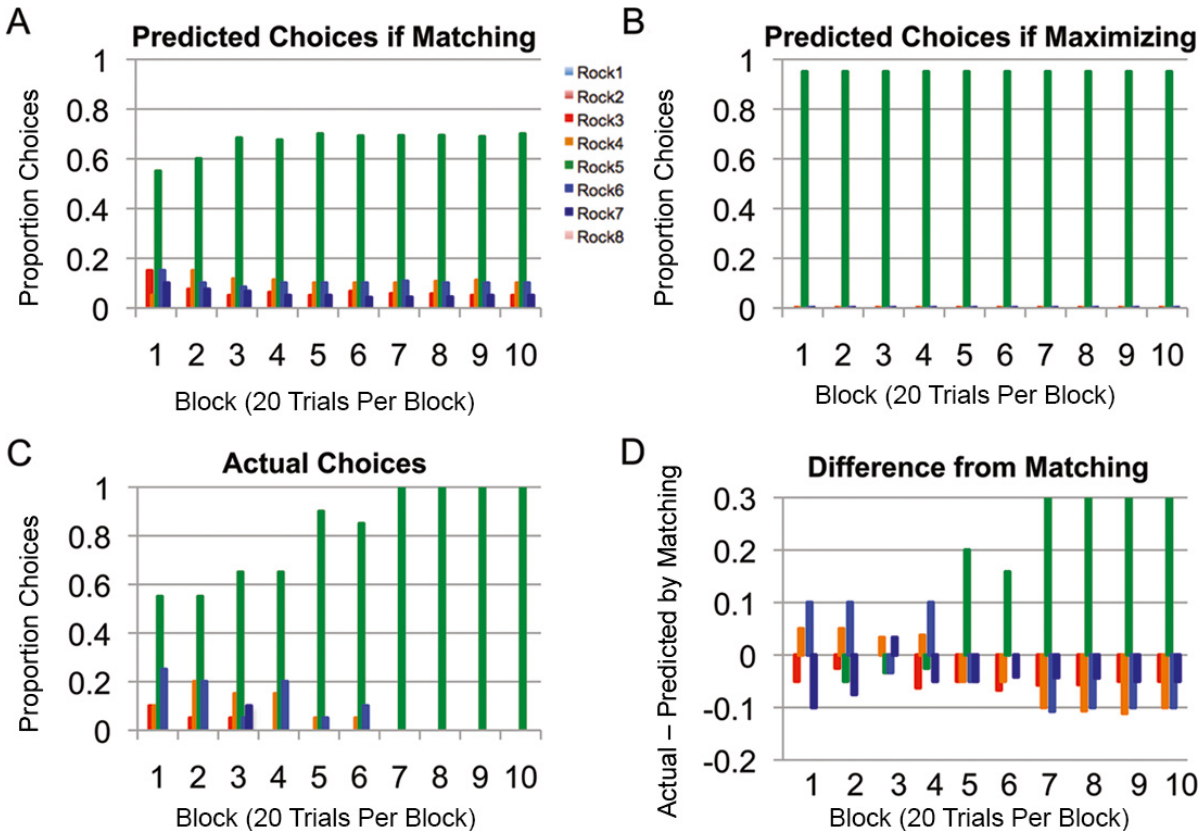

Figure 3. Modeling participant behavior. (A) The pure probability matching model predicts that participants would choose locations in proportion to the probability that those locations had been previously observed to be correct. Colored lines represent number of choices at each rock location during the task with the green line indicating choices for the peak location. (B) The pure maximizing model predicts that participants will always choose the location that had been correct the most often across the experiment. (C) We predicted that participants would use a mixture of these types of strategies. Early in the experiment, they would choose locations roughly in proportion to the probability the rocks had been correct, but later largely only choose the location that had the highest overall probability of being correct. (D) Plotting the difference between choice proportions and the expectations of probability matching, it is clear that these choices were initially largely consistent with matching (all differences near zero), but over time the participant began to disproportionately choose the location with the highest probability of being correct. Note that the depiction by trial block is for illustration purposes only—data analysis conducted as described in the text.

information for all four models are described in the Supporting Information.

In order to quantify combination model behavior, we calculated two measures that together capture the main points of interest in the model: (a) the crossover trial in which the participant's behavior deviated from being better described as matching to being better described as maximizing (referred to as the "time to crossover" for the remainder of the article) and (b) the extent to which the participant maximized at the end of the task (which captures how much more often they chose the most frequently rewarded location than would be expected by the probability matching only model).

Figure 3C depicts a hypothetical participant whose choices are most consistent with probability matching over the first 100 trials. Figure 3D reflects that this participant shows a very small difference between the proportion of choices made and the expectations given by probability matching. However, during the second half of the experiment, the participant's choices increasingly deviate from matching-like behavior toward choice behavior consistent with maximizing. That is, with time, the participant chooses the most frequently rewarded location disproportionately more often compared to the location's probability of being correct, and eventually, the participant exclusively chooses this peak location. The combination model (Figure 4) thus models behavior as initially primarily matching (i.e., the mixture is dominated by the probability matching model) and then progressively moves toward pure maximizing (i.e., the mixture is dominated by the maximizing model). If the participant in Figure 3 had started to crossover to maximizing earlier during the course of the experiment (Figure 5A), the model would capture this as a change in the time to crossover (earlier crossover; Figure 5B). If, on the other hand, the participant never quite crossed over to pure maximizing behavior (i.e., still occasionally chose rocks other than the "peak" rock, but less frequently than would be expected by probability matching; Figure 5C), the model would capture this as a change in final maximizing (Figure 5D).

Results

Thirty-one children completed 200 trials; 11 children terminated the study after 100 trials. Due to those who only completed 100 trials not having equivalent opportunity as those who completed 200 trials, they were excluded from all analyses. The final sample included 31 children (18 males; M_{age} = 7.71, SD_{age} = 1.99; 84% White) and 32 young adults (13 males; M_{age} = 20.47, SD_{age} = 1.74; 53% White [12 did not report race]).

Characterizing Responses

Consistent with our hypothesis, the majority of participants demonstrated a change in choice behavior over time. Of the 31 children and 32 adults tested, 23 (74%) children and 26 (81%) adults were best fit by the combination model. These participants exhibited primarily matching behavior at the outset of the experiment and then crossed over to primarily maximizing later in the experiment. All participants who did not crossover from matching to maximizing (26% of children, 19% of adults) were best fit by the probability matching only model,

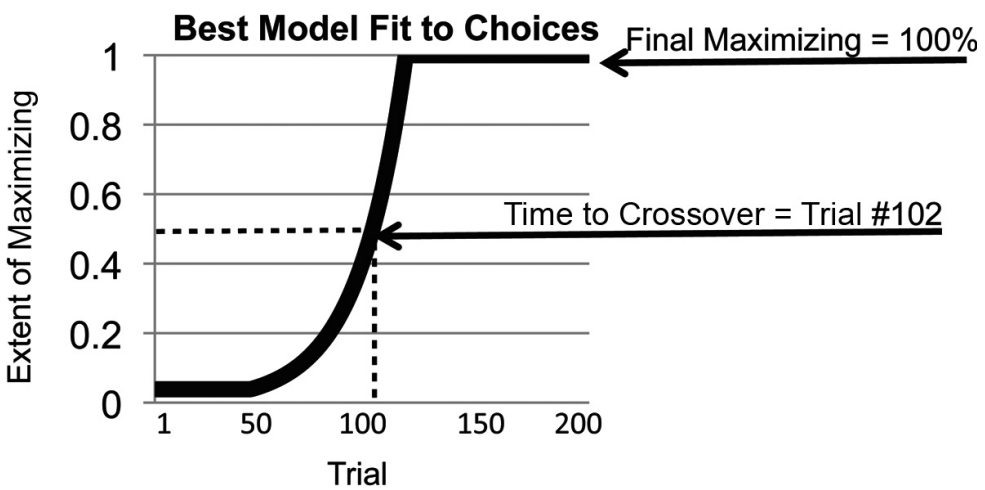

Figure 4. Model fit for example participant in Figure 3—marking the time to "crossover" and the final maximizing.

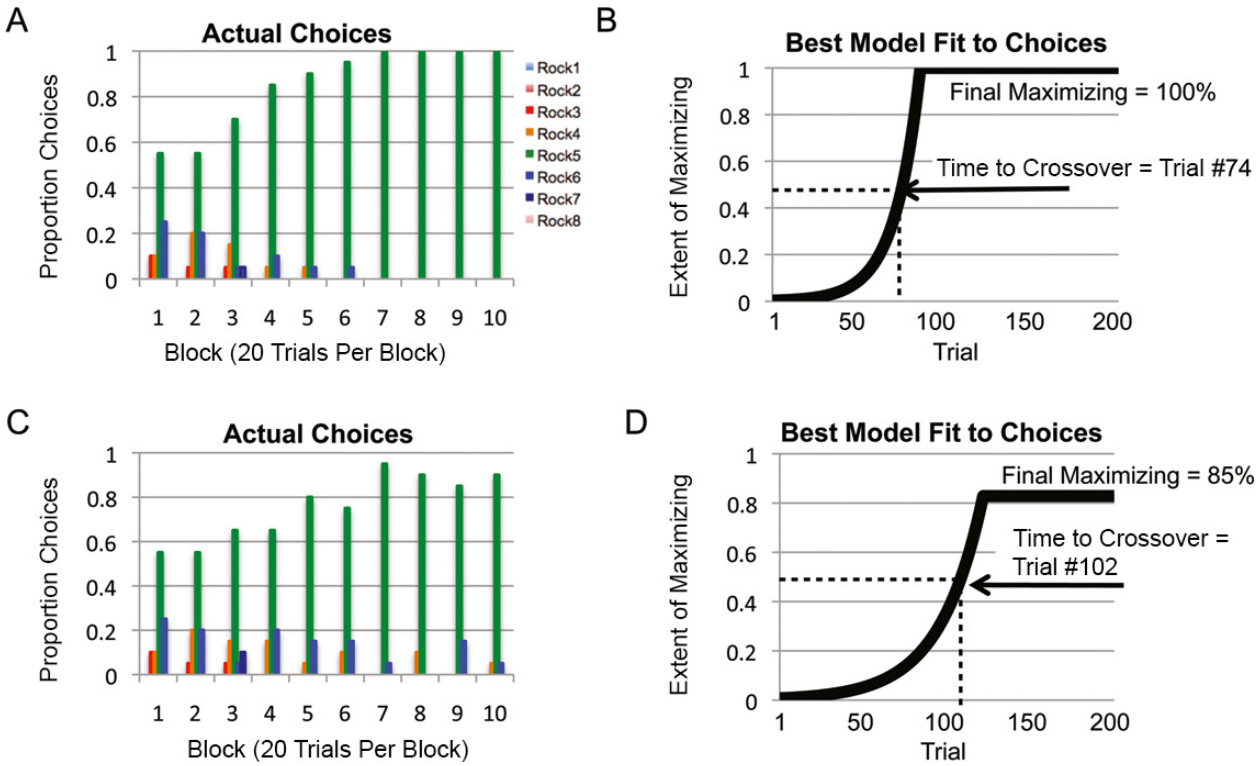

Figure 5. As in Figure 3, colored lines represent number of choices at each location during the task. The green line indicates choices for the peak location. (A) If the same participant plotted in Figure 4 had started to deviate from pure probability matching earlier in the experiment (e.g., in this case exceeding those predictions as early as Trial 60), this would be captured by the model as an earlier time to crossover, but no change in final maximizing. (B) This is seen in the combination model fit, where the time to crossover is lower than in Figure 4 (Trial 74 vs. Trial 102), but the final maximizing is identical in both (full maximizing). (C) If instead, the same participant plotted in Figure 4 did not truly maximize at the conclusion of the experiment (i.e., continued to sometimes choose options other than the peak location) but still made choices to the peak location more often than would be predicted by pure probability matching (e.g., in this case 90% of choices to the peak location rather than 70% of choices), this would be captured by the model as a change in the final maximizing. (D) This is seen in the combination model fit, where the time to crossover is the same as was seen in Figure 4, but the final maximizing is lower here (maximizing here = 85%, rather than 100% in Figure 4).

suggesting that these participants were indeed sensitive to the underlying probabilities and were choosing locations according to this information (model fits provided in Appendix).

No difference was observed between the proportion of children and adults who were best fit by the combination model, 73% versus 81%, respectively; $\chi^2(1) = 0.14$, $p = .71$. Thus, children and adults showed a similar pattern of behavior change over time. We next examined whether there were age differences in the rate of strategy change.

Time to Crossover From Matching to Maximizing

To assess timing of behavior change, we regressed the trial at which participants crossed over on age group (children vs. adults). Children crossed over later than adults ($M = 82.61$, $SD = 53.19$ vs. $M = 52.85$, $SD = 38.24$, respectively); $b = -29.76$, $R^2 = .10$, $F(1, 47) = 5.14$, $p = .03$ (Figure 6). To examine differences in time to crossover across development in children only, we used a linear regression, regressing the trial at which participants crossed over from matching to maximizing on age (continuous) for children. The relation between age and time to crossover in children was not significant, $b = 6.23$, $R^2 = .05$, $F(1, 21) = 1.06$, $p = .31$. Given our particular interest in developmental differences, additional analyses were conducted to test for age-related changes within the child sample. These additional analyses can be found in the Supporting Information, but consistently showed no relation between age and time to crossover.

Maximizing

Finally, we investigated the extent to which participants maximized at the conclusion of the

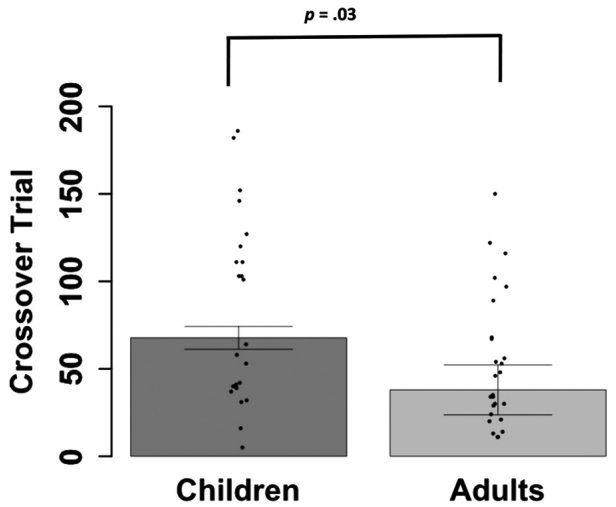

Figure 6. Trial number when participants' data were first described as maximizing. Mean trial number and raw data presented as a function of age group (includes only participants who crossed over to maximizing; children $n = 23$, adults $n = 26$). Error bars represent standard error of the mean.

experiment (again, only for participants whose data were best fit by the combination model). As discussed earlier, all participants who were best fit by the combination model crossed over from matching to maximizing at the conclusion of the experiment. However, there was some variation in maximizing. For example, some participants exclusively chose the peak location (i.e., "pure maximizing"), whereas others continued to sample other locations, albeit rarely and much less often than would be expected by a pure probability matching strategy. To test whether a later crossover to maximizing predicted lesser maximizing at the end of the experiment, we regressed maximizing on age group. Although children crossed over to maximizing later than adults did, children were maximizing to the same extent as adults at the conclusion of the experiment, a score of 1.0 would indicate exclusively choosing the peak location; children: $M = 0.92$, $SD = 0.12$ versus adults: $M = 0.95$, $SD = 0.10$; $b = 0.02$, $R^2 = .01$, $F(1, 47) = 0.59$, $p = .45$. Again, there was no relation between age and maximizing within the child age group, $b = 0.009$, $R^2 = .02$, $F(1, 21) = 0.43$, $p = .52$. Additionally, we tested whether there was a relation between the trial when participants crossed over and final maximizing. Such a relation might suggest that participants who crossover later may not be able to achieve high levels of maximizing. No such relation was observed, $r(47) = -.24$, $p = .10$, suggesting that time to crossover is unrelated to participants' tendency to reach high levels of maximizing.

Summary of Experiment 1

Most participants' choices were best modeled as a time-evolving mixture of probability matching and maximizing—beginning first with primarily matching and then crossing over to primarily maximizing. Only 26% of children and 19% of adults were best modeled as maintaining pure probability matching throughout the experiment. Furthermore, although there was similarity between adults and children at the gross level of description (i.e., the percentage best fit by combination model), children were slower than adults to change behavior. A lingering question concerned the possibility that the normal reward probability distribution used in Experiment 1 was easy or intuitive for participants to learn. Therefore, we ran an additional experiment with a non-normal reward distribution.

Experiment 2

To test the use of incoming information to direct behavior change over time using a non-normal probability distribution, we altered the reward values in two ways. First, we reduced the probability of the peak location and increased the probability of an adjacent location. Second, we moved the location of the peak option. We predicted that these changes would increase the probability that participants would maintain matching instead of crossing over to maximizing because individuals tend to show more probability matching behavior when the reward probabilities of the various alternatives are more similar to one another (Little, Brackbill, Isaacs, & Smelkinson, 1963; Moran & McCullers, 1979; Vulkan, 2000).

In Experiment 2, we also introduced an additional measure of whether participants were sensitive to the underlying reward distribution. Our goal was to obtain an explicit measure of what participants "know" about the underlying probabilistic structure by the end of the experiment.

Method

Participants

The participants in Experiment 2 included 39 children and 33 adults. None of these individuals participated in Experiment 1. Thirty-two children completed 200 trials of this task, whereas 7 children terminated the study after 100 trials and were

excluded from the subsequent analyses. The final sample included 32 children (15 males; M_{age} = 7.87, SD_{age} = 1.90; 97% White) and 33 adults (13 males; M_{age} = 20.61, SD_{age} = 1.92; 48% White).

Procedure

Experiment 2 used the same procedure as Experiment 1, except that the probabilities assigned for each rock location were (from left to right): 20%–60%–10%–5%–5%–0%–0%–0% (Figure 7; calculation of reward receipt for matching and maximizing provided in the Supporting Information). Upon completion of the experiment, adults provided estimates of the percentage of time the reward appeared at each location. Pilot testing suggested that the majority of children were unable to provide specific information related to probabilities, so we instead asked children to identify the rock behind which the most coins appeared.

Scoring and Analyses

We fit the same four models described in Experiment 1. Data from all participants were either best fit by the combination model or the probability matching only model and thus the results will be presented in the same format as in Experiment 1.

Results

Characterizing Responses

Of the 32 children and 33 adults who completed Experiment 2, 10 (31%) children and 18 (55%) adults were best fit by the combination model. All other participants were best fit by the probability matching only model. Similar to Experiment 1, there was no difference between proportion of children and adults best fit by the combination model, 31% versus 54%, respectively; $\chi^2(1) = 2.71$, $p = .10$.

Because only 10 children crossed over in Experiment 2, there was insufficient power to statistically compare children to adults or to examine the effect of age on crossover time within the child sample. However, collapsing across groups, there was no correlation between time to crossover and final maximizing, $r(26) = -.25$, $p = .20$, suggesting that time to crossover to maximizing does not affect the tendency to reach high levels of maximizing.

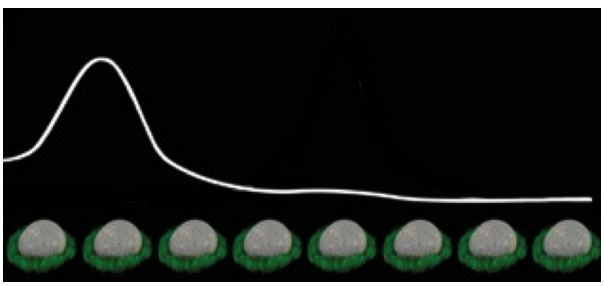

Figure 7. Distribution of rewards by location for Experiment 2. [Color figure can be viewed at wileyonlinelibrary.com]

Sensitivity to Underlying Reward Distribution

We ran a linear regression, regressing the true underlying reward distribution on participant responses and found that adults were able to successfully approximate the reward location probabilities, $b = 0.94$, $F(1, 29.73) = 991.50$, $p < .001$. There was no significant difference in the accuracy of the estimates of participants who crossed over to maximizing and those who did not, $b = -0.0000000006$, $F(1, 29.00) = .00$, $p = 1.00$, again suggesting that differences in behavior were related to choice strategy, rather than knowledge of the task statistics. All children queried (n = 21; including both those who did and did not crossover to maximizing) also correctly identified the most highly rewarded location. These results suggest that participants had a similar understanding of the distribution, regardless of whether they crossed over to maximizing.

Summary of Experiment 2

When the probabilities were altered, only 31% of children and 55% of adults crossed over from matching to maximizing. Children were particularly impacted by the probabilities, suggesting that the qualities of the information to be learned affect the strategies that adults and children are able to use.

Comparing Choice Behavior in Experiments 1 and 2

To better understand differences in behavior based on underlying probabilities, we compared participant choice behavior in Experiment 1 to participant choice behavior in Experiment 2.

Characterizing Responses

There was a significant decrease in the proportion of participants who were best fit by the combination model in Experiment 2 as compared to Experiment 1, 43% versus 78% best fit by combination model, respectively; $\chi^2(1) = 14.65$, $p < .001$. This significant difference was maintained when data from children and adults were analyzed separately, children: Experiment 1 = 74%, Experiment 2 = 31%,

$\chi^2(1) = 9.98$, $p = .002$; adults: Experiment 1 = 81%, Experiment 2 = 55%, $\chi^2(1) = 4.15$, $p = .04$.

Time to Crossover From Matching to Maximizing

To assess differences in time to crossover from matching to maximizing between Experiments 1 and 2, we regressed the trial at which participants crossed over on experimental condition. Because more adults than children crossed over to maximizing in Experiment 2, we conducted this analysis with adults only. Adults from Experiment 2 crossed over from matching to maximizing approximately 30 trials later than adults from Experiment 1 ($M = 82.67$, $SD = 54.70$ versus $M = 52.85$, $SD = 38.24$); $b = 29.82$, $R^2 = .10$, $F(1, 42) = 4.54$, $p = .04$.

Maximizing

To test whether the asymptotic level of maximizing was different in Experiment 1 versus in Experiment 2, we regressed the final proportion of maximizing on experimental condition. The linear regression indicated that adults in Experiment 2 maximized to a lesser extent than adults in Experiment 1 ($M = 0.87$, $SD = 0.17$ vs. $M = 0.95$, $SD = 0.10$, respectively) at the conclusion of the experiment, $b = -0.08$, $R^2 = .09$, $F(1, 42) = 4.36$, $p = .04$. However, there was still no correlation between time to crossover and final maximizing extent, $r(42) = -.17$, $p = .26$.

General Discussion

The aim of the present experiments was to investigate whether experience, both over the course of the experimental session and across developmental time, influences real-time behavior on a probability learning task. In Experiment 1, most participants began the task exhibiting primarily matching behavior and crossed over by the conclusion of the experiment to primarily maximizing. Experiment 2 presented participants with a different set of probabilities to learn. In this case, fewer participants demonstrated a change in strategy than in Experiment 1. This is despite the fact that participants exhibited knowledge of the probability of reward at locations in Experiment 2 after the task. Thus, participants used the outcomes of their behavior to adjust future strategies and improve their receipt of rewards over time. However, individuals required some degree of experience before they attempted this strategy change, and the amount of experience needed depended on the difficulty of the probabilistic structure.

While understanding the differences between children and adults is an area that requires more attention and will be discussed below, there were notable similarities between the choice behavior of children and adults, particularly in Experiment 1. Although children took longer to crossover, they reached similar levels of maximizing as adults by the end of the experiment. These findings suggest that children react flexibly and change behavior according to incoming information.

Differences in Behavior Across Development

The present data suggest that a simple binary classification of "matchers" and "maximizers" may miss critical information. Specifically, both children and adults changed their behavior as they gained more experience with the probability structure. However, this evolution in strategy change occurred more gradually in children than adults, suggesting that there are developmental differences in the amount of information needed before children, compared to adults, change their approach from matching to maximizing.

What remains unclear is *why* children crossover from probability matching to maximizing later than adults. Several possibilities might be considered and addressed in future research. For example, children may be less skilled than adults at learning in environments where top-down thinking (such as rejecting the hypothesis that probability matching yields greater rewards than maximizing) is needed (Lucas, Bridgers, Griffiths, & Gopnik, 2014; Newport, 1990). Instead, children may persist in matching for longer than adults because it allows them to continue sampling information about the environment (Denison et al., 2013). In certain circumstances, periods of probability matching may in actuality be valuable as a means of exploration (Seth, 2011; Shaw & Shaw, 1977), and a predisposition to explore may be particularly adaptive for children (Stephens & Krebs, 1986). Yet, children may not be as adept as adults at recognizing when exploration is advantageous or disadvantageous.

Recently, Green et al. (2010) argued that probability matching in adults does not reflect a "failure" of decision making but can in fact be an "optimal" solution given certain sensible prior beliefs about the world—prior beliefs that are violated by the probability learning task. For example, outcomes in the probability learning task are independent across time (i.e., if Location #4 was rewarded on Trial #1, it is no more and no less likely to be rewarded on Trial #2 than its base probability). Conversely, many rewards in the real world are temporally dependent (e.g., a

bush that has ripe berries on it on Monday is likely to have ripe berries again on Tuesday). A belief in temporal dependence will lead to a prolonged pattern of choices in the probability learning task that are consistent with probability matching. Under such a model, a longer period of probability matching—as is observed in children in the current experiment—would be consistent with an even stronger prior. The perspective that children may have a stronger prior than adults stands in contrast to other hypotheses posed in the literature. For example, there is some evidence that children are less reliant on their past experiences and more inclined to base choices on present evidence (Lucas et al., 2014).

Testing choice behavior under varying conditions that support the characteristics of model-free learning, such as those used in Green et al. (2010), may illuminate the role these assumptions play in choice behavior of children. Lucas et al. (2014) propose that children may either have different expectations compared to adults or be less committed to those expectations. These reasons may be relevant in the current experiments. If children are more fluid and diffuse in their beliefs than adults, they may attempt to test more hypotheses as they probability match in the present experiments, thus taking longer to commit to maximizing as their primary strategy. Therefore, while it seems unlikely that children would have a stronger prior related to temporal dependence than adults (as children have had less opportunity to observe outcomes that have temporal dependence), it could be the case that children update their beliefs less efficiently (which will also prolong the period of probability matching). This could be tested experimentally by utilizing probability learning tasks with various types of temporal dependence.

Studies examining what participants "know" about the distribution over time will also provide critical information regarding children's behavior change. For example, future studies might ask participants during each trial how confident they are that their choice will be rewarded, or assess whether participants are able to identify how likely the reward is to appear in each location. In Experiment 2, we found that all children and adults asked were able to identify the most frequently rewarded rock at the conclusion of the experiment. Therefore, knowing the peak location upon completion of the experiment does not explain why only some participants crossed over to maximizing. However, because we only asked participants at the end of the experiment, we cannot speak to whether identifying the peak location earlier in the session is related to behavior change. Further understanding the participant's subjective experience may give us other clues about individual differences in approach—for example, asking participants which approach would result in more rewards or whether they considered each approach as they were completing the task. Testing such mechanistic models will be an important next step for the future of this early research, specifically disentangling what participants "know" from what they "do." Separating these two processes may offer further insight into the nature of observed and potential developmental differences. Furthermore, we may gain insight into whether different underlying strategies are resulting in similar behaviors. For example, Bonawitz, Denison, Gopnik, and Griffiths (2014) recently provided an example of a win-stay, lose-sample shortcut that children and adults use, which leads to probability matching. Different underlying strategies may represent different stages of the learning process and further our knowledge of how children are using incoming information to direct behavior over time.

Task Considerations

Experiment 2 was designed to test the use of incoming information to direct behavior change over time using a non-normal probability distribution. Altering the underlying distribution significantly impacted performance, both in the ability of participants to execute multiple strategies as well as timing to change strategies. Future work may consider including multiple distributions in order to discern whether particular distributions are more likely to be associated with certain patterns of choice behavior. In a pilot study with adults only, we compared choices in response to the distribution in Experiment 1 with choices in response to a distribution that contained the same probabilistic values randomly assigned to each location (distribution: 10–5–5–70–0–0–10–0; n = 31). There were no statistically significant differences between performance in Experiment 1 and the pilot condition on the variables of interest (proportion of participants crossed over: p = .56, time to crossover: p = .40, extent of maximizing: p = .65). However, these findings do not rule out the possibility that children would be disproportionately adversely affected by randomly assigning probabilistic values to locations.

Another important future direction is to examine the relation between the results we have obtained in our eight alternative sequential choice task and another commonly employed type of decision-making task—the multiarmed bandit class of tasks (Gittins, 1989; Weber, 1992). While our task and

multiarmed bandit tasks have some surface similarity (e.g., multiple options where only one can be chosen on a given trial, an optimal strategy that may involve only choosing the single option with the highest reward value, etc.), there are a few critical differences that would make a future examination of a multiarmed bandit task of interest. First, in our task, the probability of reward in one location on a given trial was not independent of the other locations. Instead, on each trial a single correct location was chosen and all other locations would thus necessarily be unrewarded. In a multiarmed bandit task, whether a given location is or is not rewarded on a given trial is determined based on the location's probability independent of the other locations' outcomes (i.e., on a given trial it could be the case that all of the locations would be rewarded, none of the locations, or any particular pattern of locations). Furthermore, in our task the participant was given full feedback; they always knew which location was correct on a given trial (and in turn thus knew that the other locations were not rewarded). In a multiarmed bandit task, the participant only observes the outcome of his or her own choice. This induces what is known as an "exploration/exploitation" problem (Gittins, 1979; Gittins & Jones, 1979). In order to estimate the probability of a reward occurring at each location in a multiarmed bandit task, it is necessary to choose each location a sufficient number of times. This creates a tension between making choices in such a way as to better learn the probability of reward at each location, and making choices so as to maximize expected reward.

Conclusions

Overall, these experiments provide evidence that individuals change behavior in real time according to incoming probabilistic information. Furthermore, experience, both over time and across development, informs this ability. Individuals often encounter situations in which there is uncertainty about the relations between action and outcome. Particularly for children, who have had less experience in the world and are faced with having greater amounts of information to learn relative to adults, it is necessary to acquire effective strategies to navigate uncertainty. This research suggests that using moment-to-moment information to direct behavior helps children navigate such uncertain situations and make choices that progressively lead to advantageous outcomes as children react flexibly according to probabilities.

References

Bonawitz, E., Denison, S., Gopnik, A., & Griffiths, T. L. (2014). Win-Stay, Lose-Sample: A simple sequential algorithm for approximating Bayesian inference. *Cognitive Psychology*, *74*, 35–65. doi:10.1016/j.cogpsych.2014.06.003

Brackbill, Y., & Bravos, A. (1962). Supplementary report: The utility of correctly predicting infrequent events. *Journal of Experimental Psychology*, *64*, 648–649. doi:10.1037/h0046489

Denison, S., Bonawitz, E., Gopnik, A., & Griffiths, T. L. (2013). Rational variability in children's causal inferences: The sampling hypothesis. *Cognition*, *126*, 285–300. doi:10.1016/j.cognition.2012.10.010

Derks, P. L., & Paclisanu, M. I. (1967). Simple strategies in binary prediction by children and adults. *Journal of Experimental Psychology*, *73*, 278. doi:10.1037/h0024137

Duffy, S., Huttenlocher, J., & Crawford, L. E. (2006). Children use categories to maximize accuracy in estimation. *Developmental Science*, *9*, 597–603. doi:10.1111/j.1467-7687.2006.00538.x

Gittins, J. C. (1979). Bandit processes and dynamic allocation indices. *Journal of the Royal Statistical Society. Series B (Methodological)*, *41*, 148–177.

Gittins, J. C. (1989). *Multi-armed bandit allocation indices*. Wiley-Interscience Series in Systems and Optimization. Chichester, UK: Wiley.

Gittins, J. C., & Jones, D. M. (1979). A dynamic allocation index for the discounted multiarmed bandit problem. *Biometrika*, *66*, 561–565. doi:10.2307/2335176

Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: Causal models, Bayesian learning mechanisms, and the theory theory. *Psychological Bulletin*, *138*, 1085. doi:10.1037/a0028044

Green, C. S., Benson, C., Kersten, D., & Schrater, P. (2010). Alterations in choice behavior by manipulations of world model. *Proceedings of the National Academy of Sciences of the United States of America*, *107*, 16401–16406. doi:10.1073/pnas.1001709107

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

Little, K. B., Brackbill, Y., Isaacs, R. B., & Smelkinson, N. (1963). A further test of a general utility theory model for probability learning. *Journal of Experimental Psychology*, *66*, 107–108. doi:10.1037/h0046574

Lucas, C. G., Bridgers, S., Griffiths, T. L., & Gopnik, A. (2014). When children are better (or at least more open-minded) learners than adults: Developmental differences in learning the forms of causal relationships. *Cognition*, *131*, 284–299. doi:10.1016/j.cognition.2013.12.010

Moran, J. D., III, & McCullers, J. C. (1979). Reward and number of choices in children's probability learning: An attempt to reconcile conflicting findings. *Journal of Experimental Child Psychology*, *27*, 527–532. doi:10.1016/0022-0965(79)90041-9

Newport, E. L. (1990). Maturational constraints on language learning. *Cognitive Science*, *14*, 11–28. http://dx.doi.org.ezproxy.library.wisc.edu/10.1016/0364-0213(90)90024-Q

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926–1928.

Seth, A. K. (2011). Optimal agent-based models of action selection. In A. K. Seth, T. J. Prescot, & J. J. Bryson (Eds.), *Modeling natural action selection* (pp. 37–60). New York, NY: Cambridge University Press.

Shaw, M. L., & Shaw, P. (1977). Optimal allocation of cognitive resources to spatial locations. *Journal of Experimental Psychology: Human Perception and Performance*, *3*, 201–211. doi:10.1037/0096-1523.3.2.201

Stephens, D. W., & Krebs, J. R. (1986). *Foraging theory*. Princeton, NJ: Princeton University Press.

Vulkan, N. (2000). An economist's perspective on probability matching. *Journal of Economic Surveys*, *14*, 101–118. doi:10.1111/1467-6419.00106

Waismeyer, A., Meltzoff, A. N., & Gopnik, A. (2015). Causal learning from probabilistic events in 24-month-olds: An action measure. *Developmental Science*, *18*, 175–182. doi:10.1111/desc.12208

Weber, R. (1992). On the Gittins index for multiarmed bandits. *The Annals of Applied Probability*, *2*, 1024–1033.

Xu, F., & Garcia, V. (2008). Intuitive statistics by 8-month-old infants. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 5012–5015. doi:10.1073/pnas.0704450105

Appendix

Participants Classified as Combined Model

		Choice Strategy								
		Probability Matching			Maximizing			Combined		
		nLL	AIC	BIC	nLL	AIC	BIC	nLL	AIC	BIC
Expt 1	Children	-154.28 (8.29)	310.56 (16.59)	313.86 (16.59)	-159.99 (12.02)	321.98 (24.03)	325.28 (24.03)	-136.24 (9.59)	280.48 (19.18)	293.67 (19.18)
	Adults	-149.18 (6.39)	300.36 (12.77)	303.66 (12.77)	-143.27 (10.56)	288.54 (21.11)	291.84 (21.11)	-124.55 (8.27)	257.11 (16.53)	270.30 (16.53)
Expt 2	Children	-209.97 (12.47)	421.94 (24.95)	425.23 (24.95)	-241.73 (21.25)	485.46 (42.50)	488.76 (42.50)	-193.75 (14.49)	395.50 (28.97)	408.70 (28.97)
	Adults	-200.25 (7.01)	402.49 (14.02)	405.79 (14.02)	-216.34 (14.66)	434.69 (29.31)	437.99 (29.31)	-172.81 (10.25)	353.62 (20.50)	366.81 (20.50)

Participants not Classified as Combined Model

		Choice Strategy								
		Probability Matching			Maximizing			Combined		
		nLL	AIC	BIC	nLL	AIC	BIC	nLL	AIC	BIC
Expt 1	Children	-214.14 (23.18)	430.27 (46.36)	433.57 (46.36)	-263.09 (28.48)	528.18 (56.97)	531.48 (56.97)	-213.73 (23.18)	435.46 (46.36)	448.65 (46.36)
	Adults	-221.44 (10.98)	444.88 (21.96)	448.17 (21.96)	-279.35 (18.06)	560.69 (36.13)	563.99 (36.13)	-221.35 (11.03)	450.70 (22.07)	463.90 (22.06)
Expt 2	Children	-254.88 (14.28)	511.76 (28.57)	515.06 (28.57)	-357.53 (21.81)	717.05 (43.62)	720.35 (43.62)	-254.27 (14.34)	516.54 (28.68)	529.73 (28.68)
	Adults	-263.99 (11.25)	529.98 (22.50)	533.28 (22.50)	-354.63 (17.21)	711.26 (34.41)	714.56 (34.41)	-262.82 (11.44)	533.64 (22.88)	546.83 (22.88)

(Top) Table of model-fitting results for participants who were classified as best fit by the combined model choice strategy (i.e., a crossover time in strategy was identified in their choice behavior) in Experiments 1 and 2. (Bottom) Table of model-fitting results for participants who were not classified as best fit by the combined model choice strategy (i.e., a crossover time in strategy was identified in their choice behavior) in Experiments 1 and 2. For each of the three models, probability matching, maximizing, and combined, the negative log likelihood (nLL), the Akaike's information criterion (AIC), and the Bayesian information criterion (BIC) for the optimized fitted parameters is provided (±1 SEM). Note that for the nLL measure, larger values imply better fits and for the AIC and BIC measures, smaller values imply better fits.

Supporting Information

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

Data S1. Calculation for Matching Versus Maximizing