# It's not the treasure, it's the hunt: Children are more explorative on an explore/exploit task than adults

**Emily Sumner (sumnere@uci.edu)**
**Mark Steyvers (mark.steyvers@uci.edu)**
**Barbara W. Sarnecka (sarnecka@uci.edu)**

Department of Cognitive Sciences, University of California, Irvine,
Social and Behavioral Sciences Gateway, Irvine, California, 92697

**Abstract**

The current study investigates how children act on a standard explore-exploit bandit task relative to adults. In Experiment 1, we used child-friendly versions of the bandit task and found that children did not play in a way that maximized payout. However, children were able to identify the machines that had the highest level of payout and overwhelmingly preferred it. We also show that children's exploration is not random. For example, children selected the bandits from left to right multiple times. In Experiment 2, we had adults complete the task in Experiment 1 with different sets of instructions. When told to maximize learning, adults explored the task in much the same way that children did. Together, these results suggest that children are more interested in exploring than exploiting, and a potential explanation for this is that children are trying to learn as much about the environment as they can.

**Keywords:** cognitive development; explore-exploit; decision making

## Introduction

Imagine that it is your second day at a new job. You are standing at the coffee cart outside your office building, considering the unfamiliar menu. Yesterday you had a cappuccino and enjoyed it; today you must decide whether to get the cappuccino again like yesterday, or try the matcha green tea latte, which you might not like. This is known as an explore/exploit problem, because you must choose between exploiting a familiar option (the cappuccino) or exploiting a new one (the matcha latte). Such problems arise all the time: Do you buy the same brand of hiking boots you just wore out, or try a new style? Make the same old macaroni and cheese that you know your kids will eat, or gamble on pasta puttanesca? Re-watch that Netflix movie that you enjoyed before, or try a new one?

Researchers studying the explore/exploit problem in adults have traditionally defined a good decision as one that maximizes payout and minimizes cost. The problem is commonly operationalized in bandit tasks named after the 'one-armed bandits' (i.e., slot machines) found in casinos. In a bandit task, participants decide between two or more bandits, each of which has an unknown rate of reward. The goal of the task is to maximize return by using a combination of exploration and exploitation. Formally, the optimal strategy is to explore the different bandits just long enough to learn which one pays out best, and then switch to exploiting that one (Mehlhorn et al., 2015). Indeed, that's what most adults do. The explore-exploit problem has been widely investigated in many different contexts, including reinforcement learning (Daw, O'doherty, Dayan, Seymour, & Dolan, 2006; Wilson, Geana, White, Ludvig, & Cohen, 2014), psychiatric populations (Addicott, Pearson, Sweitzer, Barack, & Platt, 2017), and animal behavior (Beachly, Stephens, & Toyer, 1995; Chen et al., 2016; Snell-Rood, Davidowitz, & Papaj, 2011). The present studies asked, What about children?

Very few studies have investigated how children approach explore-exploit tasks, and if they approach these tasks with similar strategies as adults do. A large number of studies suggests that children, and even infants, possess intuitive statistics and a basic understanding of probability (Xu & Kushnir, 2013). Further, there is substantial evidence suggesting that children are better at maximizing statistics in scenarios whereas adults are drawn towards probability matching (Derks & Paclisanu, 1967; Hudson Kam & Newport, 2005). Taken together, this suggests that children would indeed maximize reward on this type of task, perhaps at an even faster rate than adults would. However, Blanco & Sloutsky (2018) had children complete a 4-armed bandit task on a tablet. They found that children do not maximize payout as adults typically do. Instead, they visit each bandit equally across 100 trials.

The key question in this study is that in a bandit task, in its most simplistic form, will children follow similar strategies to adults and attempt to maximize payout. If we find that children do not maximize payout, what are the reasons for their suboptimal performance? Are children viewing the task in the same way as adults?

In Experiment 1, we conduct a simplified version of the bandit task with 159 children. Critically, we designed this task to be simplistic and minimize the memory constraints which other bandit tasks possess. In Experiment 2, we conducted the same bandit task used in

Experiment 1 with adults and give different motives: to learn or to win.

## Experiment 1

One explanation for children's over-exploring in Blanco & Sloutsky (2018) could be that they simply cannot figure out which bandit pays out better. In this task, we made it easy for children to see the payouts. To do this, we showed children all of the previous results from each bandit.

We also changed the task structure so that there were three machines: One that paid off every time, one that paid off half the time, and one that never paid off.

### Methods

**Participants.** We tested 159 children between the ages of 3 and 9 (mean 5 years, 7 months; range 2 years, 11 months to 8 years, 11 months). Of those, 69 were girls, 90 were boys. Children were recruited from science museums and preschools in an urban area. An additional 11 participants were tested but excluded because they did not answer both control questions correctly. Participants were given a small toy upon sign-up (e.g., a plastic slinky).

**Procedure.** Participants were presented with three "Mystery Machines," each with a different proportion of winning (green) and losing (red) balls (see Figure 1). One box dispensed only winning balls, one dispensed only losing balls, and one alternated between winning and losing. The machines associated with different payoffs were counterbalanced across participants. Green balls contained a sticker that the child was allowed to take home; red balls contained no sticker.

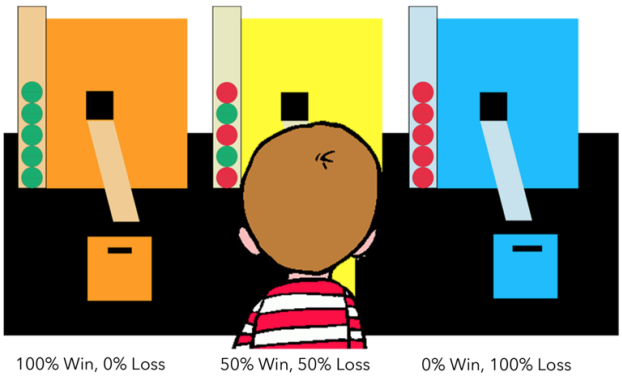


Figure 1: An illustration of the machines at the end of the study for a participant who chooses each machine equally. Green balls have stickers inside, red balls are empty. The tubes are empty at the beginning of the task.

Children were told, "These are Mystery Machines. When you put a coin in the orange box, a ball rolls down the orange slide. When you put a coin in the blue box, a ball rolls down the blue slide. When you put a coin in the yellow box, a ball rolls down the yellow slide. Balls can be green or red. A green ball, like this one, [shows green ball] has a prize that you can take home inside. See? [Opens ball and shows sticker] A red ball, like this one, [shows red ball] has no sticker in it. See? [Opens ball and shows inside.] Some machines have more green balls, some machines have more red balls."

Afterward, children were asked whether they would win any stickers if they received one green ball and one red ball. Those who were unable to correctly answer this question were excluded. Children were then given fifteen coins to put in the coin slots corresponding to the machines of their choosing. When the child put a coin in one of the three coin slots, a machine sound played, and a ball rolled down the slide of the machine corresponding to the coin slot that the child chose. Before the experiment started, a machine operator hid underneath the table at which the children were tested and crawled out from under the table once the child was seated on the other side of the table and thus could not see this person.

The machine operator used a noise maker to make the machine noise and rolled a ball down the slide each time the participant made a selection by dropping a coin in a coin slot. Empty balls were placed in the tube corresponding to the machine that they came from, in order to reduce the number of things that the participant needs to keep track of. This clearly showed the distribution of wins and losses that the participant encountered from each machine (See Figure 3).

After participants completed the task, they were asked the following questions: 1) Which machine was your favorite? 2) Why was the "xxx" colored machine your favorite? 3) Which machine has the most green balls? And 4) How do the machines work?

### Results

**Few children maximize, many explore each box equally.** Children on average pick the winning box 44.72% of the time, a value a Wilcoxon-signed rank test finds as significantly above chance ($V = 6338.5$, $p < 2.2e\text{-}16$). However, this value is still significantly below what maximizing would look like ($V = 297$, $p < 2.2e\text{-}16$). Figure 2B shows each individual child's choice.

**Children's behavior is neither random nor optimal.** To better understand children's behavior, we simulated two different strategies: a random strategy where the boxes were chosen from a uniform distribution and the optimal policy (from Steyvers, Lee, & Wagenmakers, 2009). We sampled 1000 instances from each strategy so we could compare how children acted to these datasets (See Figure 2 and Figure 3).

Children often explored the machine in a Left to Right pattern as if they were reading a book (see Figure 4) and therefore deviated from a pure random strategy of picking a

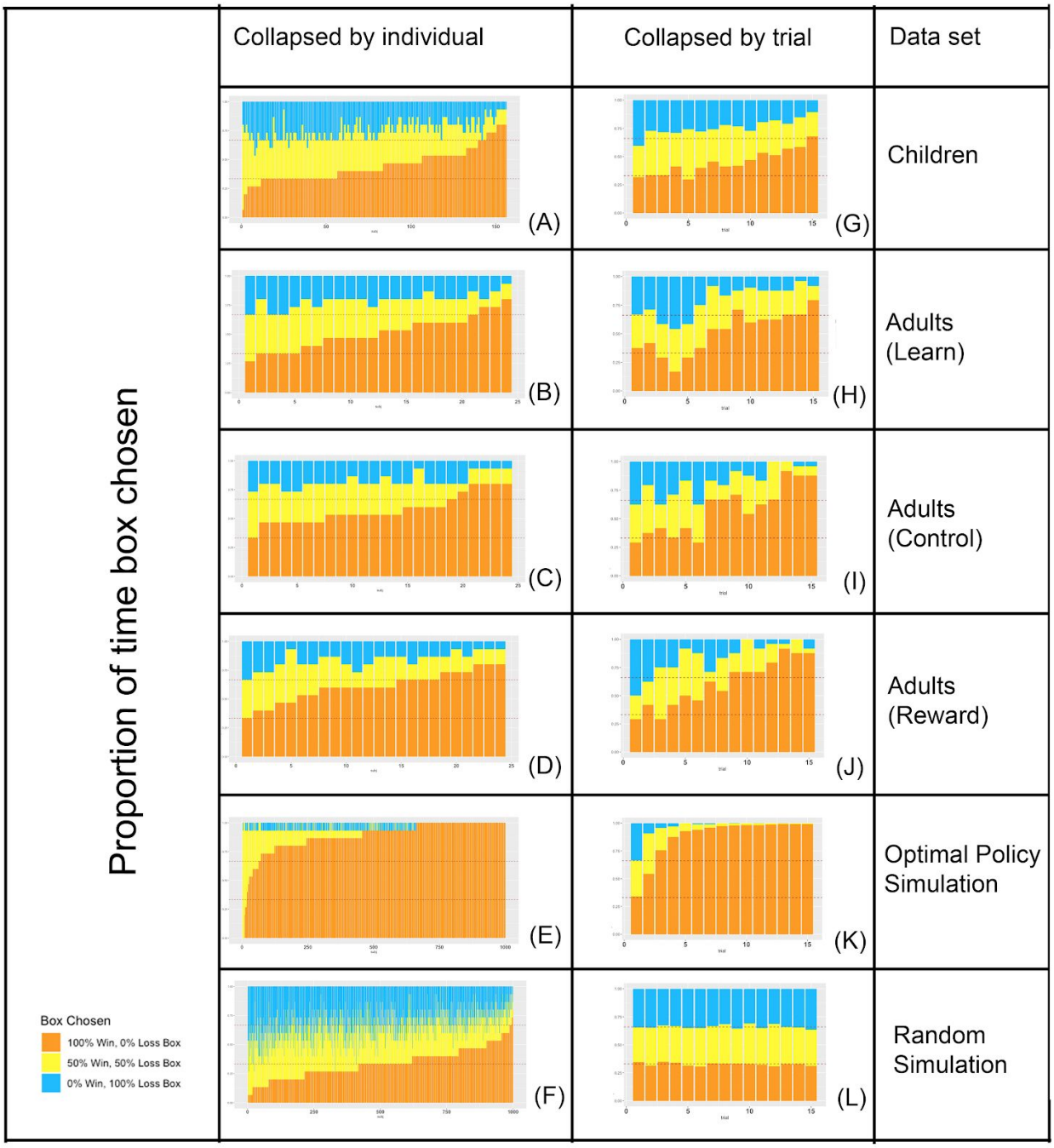


Figure 2: The proportion of choices for each box. The dotted red line indicates chance (.3333 and .6666). Orange indicates the proportion of time choosing the 100% win box, yellow indicates the proportion of time choosing the 50% win box, and blue indicates the proportion of time choosing the 0% win box. Adults in the control, reward, and the optimal policy data chose the winning box significantly more than children. Child data was collected as part of Experiment 1. The adult data were collected as part of Experiment 2. Optimal Policy data was simulated following the optimal policy in Steyvers, Lee, & Wagenmakers (2009). (A-F) Each line is a participant or simulated data point from a particular group. The red dotted line indicates chance (.3333) and (.666). Participants are ordered in increasing levels of picking the winning box. (G-H) Each line represents the proportion of participants that choose each box on a given trial.

machine at chance regardless of the previous machine chosen. For example, they would choose the orange machine, the yellow machine, and then the blue machine. Some children continued with this strategy throughout the entire task. A chi-square test of independence was performed to examine the relationship between the frequency of children going from left to right, and the frequency of this occurring in our randomly generated data set. This relationship was significant. $X^2(5, N = 1159) = 105.52$, $p = < 2.2e\text{-}16$. This suggests that children moving from left to right between boxes would not be plausible under random chance. Children could be acting in this way to strategically explore their environment, and are not necessarily acting purely randomly.

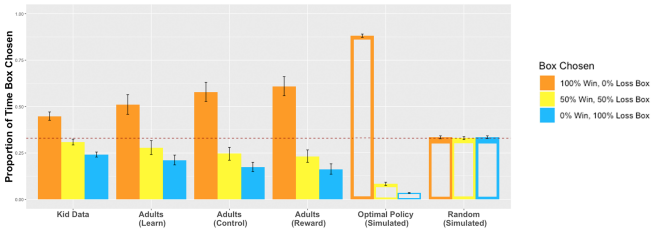


Figure 3: The proportion of times choosing each box. The dotted red line indicates chance (.3333). The error bars indicate 95% confidence intervals. Adults in the control, reward, and the optimal policy data chose the winning box significantly more than children. Child data was collected as part of Experiment 1. The adult data were collected as part of Experiment 2. Optimal Policy data was simulated following the optimal policy in Steyvers, Lee, & Wagenmakers (2009).

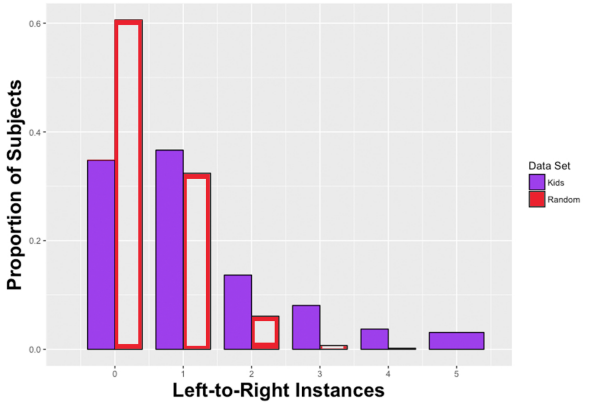


Figure 4: Histogram comparing the Left-to-Right instances made by children (purple columns) and the data set we generated where the choice of machine was random (red-outlined columns).

**Children prefer the 100% win box.** 128 children (83%) said their favorite box was the one that had 100% winning balls. 15 children (9.7%) prefer the box which alternated between winning and losing balls, 8 children (5.2%) preferred the box that always dispensed losing balls, 5 children (3.2%) said they liked all of the boxes, 2 children (1.3%) said they liked two of the boxes, and 1 child (.06%) said they didn't like any of the boxes (see Figure 5). We performed a proportion test to see if this number was significantly above chance (.3333). We found that this was unlikely due to chance $X^2(1, N = 159) = 159.2$, $p < 2.2e\text{-}16$. Of the 128 children who said their favorite box was the one that had the highest payout, we asked the children why the box they chose was their favorite, 105 children (97.2%) said it was because it gave the most green balls. The other explanations included that they liked the color of the box, or they gave an unrelated answer such as, "because rainbow."

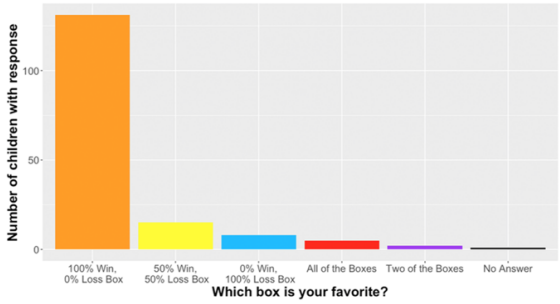


Figure 5: Children's responses to, "Which box is your favorite?"

**Age is not related to the proportion of max choices.** We looked for a correlation between age and number of max choices made. We found that age and number of times choosing the max box were not correlated. We found evidence in favor of the null ($r(159) = .086$, $BF_{01} = 5.17$).

## Experiment 2

In Experiment 1, most children did not maximize stickers. However, at the end of the task, 83% of the participants stated that their favorite machine was the one that always won. This suggests that perhaps children knew, and preferred the winning machine, but chose to explore anyways. Additionally, we found that children were making their choices in a sequential pattern that would not be predicted by chance. While it is plausible that children have not learned the distributions yet and that is why they are exploring more, in Experiment 1, children were able to identify the machine or side that had the highest payout. Children might be approaching this task with a different motive than adults.

As adults, we have more experiences with gaming and decision making. We have learned that often times, the best thing to do is maximize reward. Maybe children, who have

less experience with decision-making tasks of this nature, could just be trying to pick up as much information about the environment as possible. This would mean exploring even when you are fairly confident about the stability of an environment. Perhaps this difference in goals is what can account for differences in behavior.

To test this hypothesis, we had adults complete the protocol outlined in Experiment 1 with one of three sets of instructions: a control scenario, a reward scenario, and a learning scenario. In the control scenario, adults were told exactly what the children were told. In the reward scenario, adults were told that they would be evaluated on how many stickers they won. In the learning scenario, adults were told they would be evaluated based on how well they learned the different distributions of the three machines.

## Methods

**Participants.** We recruited and tested 72 adults from the University of California, Irvine SONA system. Our participants were primarily female (55 females, 17 males) and between the ages of 18-21 (57). 15 other participants were between the ages of 22-30. Participants were compensated with course credit.

**Procedure.** We followed the procedure outlined in Experiment 1. Adults were randomly assigned to one of three scenarios: Control, Reward, or Learning. In the Control scenario, adults were told exactly what the children were told in Experiment 1. In the Reward scenario, adults were told that they would be evaluated on how many stickers they won. In the Learning scenario, adults were told they would be evaluated based on how well they learned the different distributions of the three machines.

## Results

**Comparing adult data to children's data** Adults in the Control & Reward conditions picked the winning box significantly more than the children and the adults in the Learn condition (See Figure 2, Figure 3, & Table 1).

Using an ANOVA, we found that there was a significant relationship between the instruction condition and the number of times the winning box was chosen ($F(3,229) = 15.41$, $p < .001$, $\eta = .168$). When comparing the children data, adult data, random strategy, and optimal policy, using an ANOVA there is an even larger effect and a significant relationship between condition and number of times the winning box was chosen ($F(5, 2227) = 1570$ $p < .001$, $\eta = .779$). We found that children picked the winning box significantly less than adults in the Control condition & Reward condition, but not the Learning condition.

Table 1: Pairwise Comparison using t-tests with pooled standard deviation. t-value (p-value)

| | Control | Children | Learn |
|---|---|---|---|
| Children | -4.274 (1.0e-04) | --- | --- |
| Learn | -1.652 (0.197) | 2.09 ( 0.109) | --- |
| Reward | 10.519 (0.449) | 5.273 (8.7e-07) | 2.409 (0.064) |

**Children switch between boxes more than adults do.** A way of quantifying exploration is through looking at the proportion of switch trials participants made. If a participant is exploring, they would have a high proportion of switch trials. A participant who never makes the same choice twice in a row would have a switch proportion of 1. A participant who always chooses the same box would have a switch proportion of 0. On average, child participants had a switch proportion of .8051, whereas adults had an average switch proportion of 0.5962. Conditions that adults were in did not influence their switching behavior. A Wilcoxon rank sum test with continuity correction showed that children switched significantly more than adults did ($W = 2400$, $p = 1.527e-12$).

**Trial level analysis.** When looking at the data at the trial level (See Fig 2G-L), children and adults do choose the winning box more frequently near the end of the task. We did a logistic regression looking at trials as a predictor of choosing the winning box. For children, we found that trial number was an indicator of choosing the winning box ($\beta = 0.094$, $p < 2e-16$). However, for adults, the coefficient was higher ($\beta = 0.181$, $p < 2e-16$). This shows that while children are more likely to chose the maximal option as the task goes on, the change at a much slower rate than adults.

## Discussion

In this paper, we presented two bandit experiments had no memory constraints. The first was with children, who did not play in a way that maximized payout and explored more than would be optimal. Uniquely, our paper showed that children were able to identify the machines that had the highest level of payout and overwhelmingly preferred the bandit with the highest payout. We also show that children's exploration is not random. For example, children moved across the bandits from left to right over and over again, as if they were reading a book.

In Experiment 2, we instructed adults to either maximize payout or learn the distributions of the 3-armed-bandit task. Experiment 2 showed that adults maximized payout, but when they are asked to maximize learning, they explore more -- like children.

Together, these results suggest that children are more interested in exploring than exploiting, and a potential explanation for this is that children are trying to learn as much about the environment as they can. There are several possible explanations for our findings.

**Explanation 1: Children don't maximize payout because they don't know which machine pays out the best.** One potential explanation for our data is that children spend longer than adults exploring the environment of the game because it takes children longer to figure out which machine has the best payout. This is plausible, given that children have much poorer working memory than adults do (Gathercole, Pickering, Ambridge, & Wearing, 2004), as well as less experience with this type of task.

But the children in our study did know which machine gave the most stickers -- at least, they knew this by the time they finished playing. And children overwhelmingly preferred the machine with the best payout.

Moreover, despite their poorer working memory, there are some domains of learning where children come to the correct answer faster than adults do. In language learning, for example, 6-year-old children outperform adults by maximizing probability whereas adults tend to match probability (Hudson Kam & Newport, 2005). Children also outperform adults in a simple probability guessing game (Derks & Paclisanu, 1967).

**Explanation 2: Children don't maximize payout because they would rather explore the game environment.** In any explore/exploit task, participants must explore to find the resources before they can shift to exploiting those resources. But it is possible that the shift from exploring to exploiting is not only seen over the course of an individual task, but over many timescales. Just as all participants explore in the early stages of a task, perhaps people explore more (in general, across domains) in the early years of life -- that is, in childhood. the time scale of one human life. Perhaps children are more 'exploring' than adults in general, meaning that they seek information about the environment in a broad sense rather than in just the narrow sense needed to maximize immediate payout (in this case, stickers).

According to this explanation, children sacrifice payout in order to get more information. But presumably, if children could get that information, either way, they would still want to maximize payout. And indeed, in bandit tasks where children are given all of the information that they would have gained from each of the different choices, they do maximize payout (Plate, Fulvio, Shutts, Green, & Pollak, 2018; Starling, Reeder, & Aslin, 2018).

We hypothesize that the optimal time to shift from exploring to exploiting depends on (A) how well you know the environment, and (B) how likely it is that the environment will change. When you don't know the environment well and/or the environment is likely to change, then more exploring is beneficial because it provides more information about all aspects of the environment (addressing Problem A) and it provides information that may be helpful if something in the environment changes (Problem B.) We hypothesize that children typically are in a situation where (A) is low and (B) is high, so they naturally explore, whereas adults know the environment better and also have fewer years left ahead of them, meaning that the amount of change they must prepare for is lower.

Our results are consistent with the idea that children develop flexible knowledge through exploration and broader search (Gopnik et al., 2017). From a child's point of view, the world is constantly changing. It makes sense to prioritize gathering data rather than maximizing immediate payouts. As our Experiment 2 showed, adults do the same when they are told to focus on learning, rather than on immediate rewards.

## Acknowledgments

We thank the University of California, Irvine Graduate Division and Undergraduate Research Opportunity Program for supporting this project. We thank the Discovery Cube of Orange County & the Montessori Schools of Irvine for allowing us to collect data; Chelsea Parlett Pelleritti & Mac Strelioff for useful discussion; Fatima Pineda, Tina Singh, Mikaya Hand, Kelly Fogarty, Kelsy Chou, Julissa Navas, Hasmik Mehrabyan, Amanda Jamison, and Eden Harder for assistance with data collection.

## References

Addicott, M. A., Pearson, J. M., Sweitzer, M. M., Barack, D. L., & Platt, M. L. (2017). A primer on foraging and the explore/exploit trade-off for psychiatry research. Neuropsychopharmacology, 42(10).

Beachly, W. M., Stephens, D. W., & Toyer, K. B. (1995). On the economics of sit-and-wait foraging: site selection and assessment. Behavioral Ecology, 6(3), 258–268.

Chen, W., Koide, R. T., Adams, T. S., DeForest, J. L., Cheng, L., & Eissenstat, D. M. (2016). Root morphology and mycorrhizal symbioses together shape nutrient foraging strategies of temperate trees. Proceedings of the National Academy of Sciences, 113(31), 8741–8746.

Daw, N. D., O'doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. Nature, 441(7095), 876.

Derks, P. L., & Paclisanu, M. I. (1967). Simple strategies in binary prediction by children and adults. Journal of Experimental Psychology, 73(2), 278–285. http://dx.doi.org/10.1037/h0024137

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. Developmental Psychology, 40(2), 177.

Gopnik, A., O'Grady, S., Lucas, C. G., Griffiths, T. L., Wente, A., Bridgers, S., … Dahl, R. E. (2017). Changes in cognitive flexibility and hypothesis search across human life history from childhood to adolescence to adulthood. Proceedings of the National Academy of Sciences, 114(30), 7892–7899. https://doi.org/10.1073/pnas.1700811114

Hudson Kam, C. L., & Newport, E. L. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. Language Learning and Development, 1(2), 151–195.

Mehlhorn, K., Newell, B. R., Todd, P. M., Lee, M. D., Morgan, K., Braithwaite, V. A., … Gonzalez, C. (2015). Unpacking the exploration–exploitation tradeoff: A synthesis of human and animal literatures. Decision, 2(3), 191.

Plate, R. C., Fulvio, J. M., Shutts, K., Green, C. S., & Pollak, S. D. (2018). Probability Learning: Changes in Behavior Across Time and Development. Child Development, 89(1), 205–218. https://doi.org/10.1111/cdev.12718

Snell-Rood, E. C., Davidowitz, G., & Papaj, D. R. (2011). Reproductive tradeoffs of learning in a butterfly. Behavioral Ecology, 22(2), 291–302.

Starling, S. J., Reeder, P. A., & Aslin, R. N. (2018). Probability learning in an uncertain world: How children adjust to changing contingencies. Cognitive Development, 48, 105–116.

Wilson, R. C., Geana, A., White, J. M., Ludvig, E. A., & Cohen, J. D. (2014). Humans use directed and random exploration to solve the explore–exploit dilemma. Journal of Experimental Psychology: General, 143(6), 2074.

Xu, F., & Kushnir, T. (2013). Infants are rational constructivist learners. Current Directions in Psychological Science, 22(1), 28–32.