Preschoolers Are Sensitive to Their Performance Over Time

Julia A. Leonard[1] (jlnrd@sas.upenn.edu), Julia Sandler[1] (sandlerj@sas.upenn.edu), Amanda Nerenberg[1] (amandarn@sas.upenn.edu), Aidan Rubio[1] (aidanrub@sas.upenn.edu), Laura E. Schulz[2] (lschulz@mit.edu), Allyson P. Mackey[1] (mackeya@upenn.edu)

[1]Department of Psychology, University of Pennsylvania

[2]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology

Abstract

Tracking one's performance over time is essential to efficient self-guided learning but it is not clear whether young children can accurately monitor their past performance. Here, we looked at whether 4-6-year-olds can use the trajectory of their past performance to allocate future resources. Across four experiments (*N* = 274), we found that children were sensitive to their rate of change in past performance: Children assigned to a condition in which they got better over time were more likely to take on challenges and teach others than children in conditions where they got worse or stayed the same. Furthermore, children privileged their rate and direction of change more than their total or final score. These results suggest that young children monitor their rate of improvement and can use this information to guide their future efforts.

Keywords: Challenge seeking, Confidence, Ability, Effort, Reasoning

Introduction

Young children's learning is often self-guided and not formally supervised, leading to many situations in which they get to choose how to challenge themselves. They decide whether to make bigger or smaller block towers or read easier or harder books. Yet little is known about how young children make these decisions that are so critical to their growth.

As adults, we often consider our past performance when making decisions about how to allocate resources or take on challenges. When adults evaluate their own learning, they allocate more practice time to unlearned, difficult items; thus adults' judgments of learning are correlated with their actual performance (Dunlosky & Nelson, 1992) and academic achievement (Young & Fry, 2008). Understanding one's own learning curve across domains and tasks is advantageous for maximizing efficient learning, preventing people from investing time in an area where they are unlikely to see gains (Nelson & Leonesio, 1988) or from focusing on tasks where their attainment is already quite high (Metcalfe & Kornell, 2005). However, it remains an open question whether young children, like adults, learn from changes in their performance over time to allocate future resources.

One reason to believe children might be capable of using their past performance to calibrate the effort they expend in the future is that across a variety of domains, they learn rationally from data —that is, they make inferences about the world by integrating their prior beliefs with new evidence (see Schulz, 2012; Tenenbaum et al., 2011). This is true not only when children are learning about the world around them (e.g., Bonawitz et al., 2012; Stahl & Feigenson, 2015) but also when they are trying to learn about their own abilities (Leonard et al., 2017; Lucca et al., 2020). For example, four and five-year-olds are much more likely to indicate that they are "bad at solving puzzles" after experiencing failure than at baseline (Smiley & Dweck, 1994) and are more likely to have negative reactions and want to switch to a different task after failure (Stipek et al., 1992). Even very young children seem to have some metacognitive awareness of their own abilities: 20-month-olds ask for help when they are uncertain (Goupil et al., 2016) and four and five-year-olds proactively select evidence that will be easier for them to discriminate (Siegel et al., 2014). However, while previous work suggests that children can consider both their immediate past failures and the probability of information gain given uncertainty, it does not address the question of whether children are sensitive to changes in their performance over time.

Moreover, considerable research suggests that children are optimistic about their abilities and often fail to learn from their past successes and failures. Four and five-year-olds are over-confident about their memory span, and, after doing poorly on a memory test, fail to allocate more study time when presented with the same task again (Flavell et al., 1970). Indeed, even children as old as eight fail to modify their study time based on task difficulty (Metcalfe & Finn, 2013). Similarly, relative to older children, preschoolers give higher estimates of their ability to perform motor tasks after initial failure (Schneider, 1998; Stipek & Hoffman, 1980). One possibility is that children may be able to track their current uncertainty but fail to track past outcomes when predicting future performance (Parsons & Ruble, 1977). Another possibility is that young children engage in "wishful thinking" and don't differentiate their wishes from their expectations (Schneider, 1998).

In short, it is not clear to what extent children use changes in their past performance to predict their future performance and calibrate their effort accordingly. In the current paper, we explore young children's sensitivity to trajectories of past performance. Across four experiments (three pre-registered), we randomly assigned children to conditions in which they got better, worse, or stayed the same at a task over time and looked at how this impacted their subsequent behavior. We predicted that children assigned to conditions where they got better over time would be more likely to choose to do a more challenging version of the task (Exp. 1- 2) and to teach someone else about the task (Exp. 1a-c), than children assigned to a condition where they got worse or stayed the same at the task over time.

Experiment 1a

Methods

Participants Participants were recruited at an urban children's museum and tested individually in quiet testing rooms off of the museum floor. Fifty-one 4-5-year-old children were recruited for the study, but only 48 were included in the data analysis (mean age: 59.49 months; range: 48 - 71 months) due to experimental error (n = 3). Children were randomly assigned to one of two conditions: Increasing Performance or Decreasing Performance; ages were matched across conditions, (b = 0.25, 95% CI [-3.94, 4.37][1]). This initial exploratory experiment was run with 24 subjects/ condition (Increasing: 8 F; Decreasing: 12 F). Full data was only available on 43 children (21 Increasing, 22 Decreasing), due to 5 children not answering the puppet question. All research in Experiments 1-4 were approved by the Institutional Review Board and conducted with the informed consent of parents.

Procedure Children were introduced to a ball throwing game. They were instructed to stand on an X mark behind a taped line on the floor and try to throw 5 balls into a basket 90 cm away on the floor. This served as a calibration trial - if children got 0-1 balls in, the basket was moved closer to the child for the remaining trials (60 cm away) and if they got 4-5 balls in, the basket was moved farther away for the remaining trials (120 cm away). If the child got 2-3 balls in, the basket remained 90 cm away for all remaining trials. This enabled us to control for individual differences in children's ball-throwing ability.

Next the experimenter introduced the child to the actual game, instructing children to try to throw as many balls into the basket before the timer (a buzzer secretly controlled by the experimenter) goes off. Once the child got the set amount of balls in the basket (see Figure 1 for conditions), the experimenter would ring the buzzer letting children know that the trial was over. This form of surreptitious task structure has been used in other studies (Rhodes & Brickman, 2008). The experimenter repeated this for four trials, moving the filled basket off to the side and replacing it with a new one each time. At the end of the trials the experimenter lined up the four baskets (with their balls in them) in front of the child and said out loud how many balls the child got in on each round.

Next the child had to make two choices, both meant to provide indices of the child's sense of their own confidence at the task. First, they had to choose how they wanted to play the game for the last (5th) trial. The experimenter put out two baskets: one that was closer to the child and wider (60 cm away) and the other that was tall and farther from the child (105 cm away). Children were told that they could play the game one more time with the basket "that is bigger and closer to you so it will be easier" or with the basket "that is taller and farther away from you, so it will be a challenge". Next children could choose whether they wanted to learn from a knowledgeable puppet (i.e., felt relatively less confident) or teach an ignorant puppet (felt relatively more confident) about the game. The order of these two outcome measures was fixed.

A combination of these two measures created the outcome variable of interest: the self-perceived confidence score, which ranged from 0-2. For both the basket and puppet measures, children got a score of 0 if they chose to play with the easier bucket/ teacher puppet, and 1 if they chose to play with the harder bucket/ learner puppet option. These two scores were added together to create the confidence score. This measure encompasses the participant's decision about both of their actions and social interactions as a consequence of their past performance on the task.

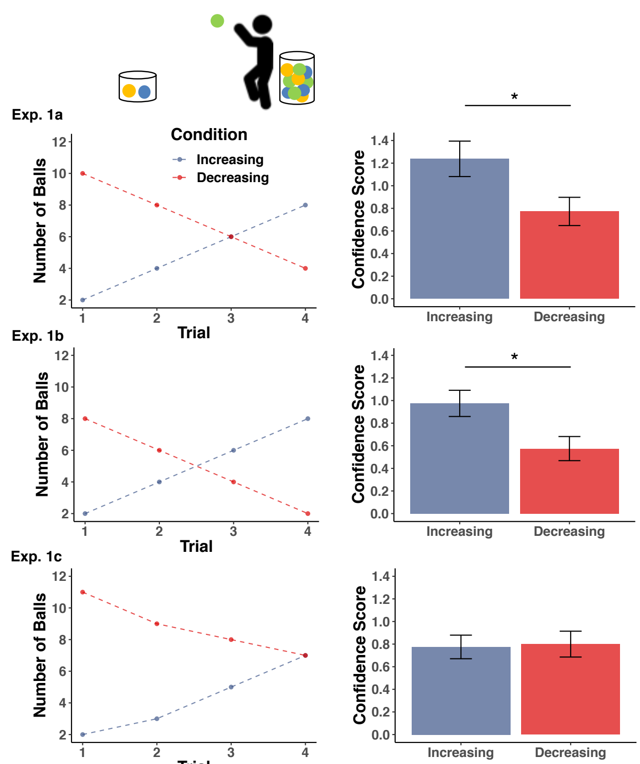

Figure 1. Experimental learning curves and results for Exp. 1a-c. On each trial, children tried to get balls into a basket. On the left are graphical representations of the learning curves (number of balls in basket on each trial) per Exp. On the right are bar charts (mean and SE) of children's confidence score by condition. *p < .05

[1] All reported CIs are 95% confidence regions estimated through a basic non-parametric bootstrap of the data using 10,000 samples

Results & Discussion

Children in the Increasing condition had significantly higher confidence scores than children in the Decreasing condition (Increasing mean: 1.24, Decreasing mean: 0.77; $W = 311$, $p = .04$, $r = -.32$). To look at whether this difference was driven by the basket choice or the puppet choice, we ran two post-hoc chi-square tests. There was a trend for children in the Decreasing condition to choose the easier basket more often than children in the Increasing condition ($\chi^2(1, n=48) = 3.14$, $p = .08$). There was no difference in puppet choice by condition ($\chi^2(1, n=43) = 1.26$, $p = .26$). Age did not relate to the confidence score ($\rho(41) = .16$, $p = .31$), puppet choice ($t(41) = -0.66$, $p = .51$), basket choice ($t(46) = -0.24$, $p = .81$), or number of calibration throws in ($\rho(44) = .01$, $p = .97$).

A number of control analyses were run to make sure individual differences in children's performance did not drive the difference of confidence score by condition. Children's actual skill at ball throwing, indexed by the number of calibration throws in, did not differ by condition ($t(46) = .48$, $p = .63$) or relate to their confidence score ($\rho(41) = .04$, $p = .81$). The total number of balls children missed (didn't get into the basket) also did not differ by condition ($t(46) = 0.85$, $p = .40$) or correlate with confidence scores ($\rho(41) = .07$, $p = .65$). As expected, children in the Decreasing condition got faster with each trial (because they were getting less balls in as time went on) than children in the Increasing group, resulting in a significant interaction of trial by condition on time per trial ($F(3, 140) = 5.93$, $p = .001$). However, individual differences in time per trial from trials 1 to 4 did not relate to confidence scores within condition (Increasing: $\rho(15) = .25$, $p = .30$; Decreasing: $\rho(12) = .06$, $p = .81$). Thus, individual's sensitivity to the manipulation, as indexed by speeding up or slowing down, did not impact their confidence score.

Exp. 1a established that children are sensitive to their performance over time and use this information to calibrate future actions. However, this study was exploratory in nature, with the goal of establishing that children could understand the paradigm and were sensitive to the manipulation. This study also helped establish an effect size, enabling us to run a power analysis for a replication study. In Exp. 1b, we aimed to replicate our effect with an adequately powered sample size and run a more controlled experiment by matching the total score across the two conditions (in Exp. 1b, both conditions have 20 balls total).

Experiment 1b

Methods

Participants & Procedure Exp. 1b was pre-registered on the Open Science Framework (OSF; https://osf.io/tn8br). Participants were recruited in the same fashion as in Exp. 1. Ninety-two 4-5-year-old children were recruited for the study, but only 80 were included in the data analysis (mean age: 59.05 months; range: 48 - 70 months) due to parental interference ($n = 4$), children voluntarily withdrawing ($n = 5$), or experimental error ($n = 3$). Children were randomly assigned to an Increasing or Decreasing condition; ages were matched across conditions ($b = 0.20$, 95% CI [-2.64, 3.04]). We ran a simulated power analysis using the data from Exp. 1a to determine a sample size for Exp. 1b. This analysis revealed we would need to collect data on 40 children per condition to find a large difference between conditions (power = .80). We collected data on 80 children (40/ condition; Increasing: 20F, Decreasing: 20F) in Exp. 1b. The procedure of Exp. 1b was the same as Exp. 1a except in the Decreasing condition, children got 8, 6, 4, and then 2 balls in.

Results & Discussion

Again, as predicted, children in the Increasing condition had higher confidence scores than children in the Decreasing condition (Increasing mean: 0.98, Decreasing mean: 0.56; $W = 1037$, $p = .01$, $r = -.28$). To examine whether this difference was driven by the basket choice or the puppet choice, two post-hoc chi square tests were run. There was no difference of basket choice by condition ($\chi^2(1, n=80) = 1.1$, $p = .30$). However, children in the Increasing condition were more likely to choose to teach a puppet and children in the Decreasing condition were more likely to choose to learn from a puppet ($\chi^2(1, n=80) = 6.11$, $p = .01$). Again, age did not relate to confidence scores ($\rho(78) = -.09$, $p = .41$).

The same control analyses from Exp. 1a were run in Exp. 1b to make sure individual differences in children's performance did not drive the difference in confidence score by condition. Children's number of calibration throws in did not differ by condition ($t(78) = 1.12$, $p = .27$) or relate to their confidence score ($\rho(78) = -.01$, $p = .91$). The total number of balls children missed did not differ by condition ($t(78) = 0.15$, $p = .88$) or correlate with confidence scores ($\rho(78) = -.09$, $p = .41$). Again, there was a significant interaction of trial by condition on time per trial, with children in the Decreasing condition getting faster from trials 1-4 than children in the Increasing condition ($F(3, 304) = 11.75$, $p < .001$). As in Exp. 1a, individual's sensitivity to the manipulation, indexed by the difference in time per trial from 1- 4, did not relate to confidence scores within conditions (Increasing: $\rho(38) = -1.04$, $p = .31$; Decreasing: $\rho(38) = .81$, $p = .45$).

Exp. 1b replicated the results of Exp. 1a: Children displayed more confidence when they improved on a game rather than got worse over time. While this provides initial evidence that children are sensitive to their rate of past performance, it does not rule out the possibility that children simply use their final performance to calibrate their future actions. Children in the Increasing condition, who end with eight balls in the basket, may feel more positive than children in the Decreasing condition, who end with two balls, irrespective of their learning over time. To rule out the possibility that children are simply relying on their final score, rather than learning over time, we ran a third experiment where the Increasing and Decreasing conditions had a matched final score.

Experiment 1c

Methods

Participants & Procedure Exp. 1c was pre-registered on the OSF (https://osf.io/dgewp). Participants were recruited in the same fashion as in Exp. 1a & b. Eighty-nine 4-5-year-old children were recruited for the study, but only 80 (Increasing: 20F; Decreasing: 23F,) were included in the data analysis (mean age: 59.05 months; range: 48 - 70 months) due to parental interference (n = 2), children voluntarily withdrawing (n = 5), or experimental error (n = 2). Children were randomly assigned to an Increasing or Decreasing condition; ages were matched across conditions, b = 1.44, 95% CI [-1.65, 4.58]). The procedure of Exp. 1c was the same as Exp. 1a & b except for the number of balls in per trial (see Figure 1).

Results & Discussion

Unlike Exps. 1a and b, in Exp. 1c children's confidence scores did not differ by condition (Increasing mean: 0.78, Decreasing mean: 0.80; W = 791.5 p = .93). Similarly, children did not differ in their basket choice (χ^2(1, n=80) = 0.26, p = .61) or puppet choice by condition (χ^2(1, n=80) = 0.45, p = .50). Again, age did not significantly relate to the confidence score (ρ(78) = -.02, p = .88).

When children's performance on the last trial was matched, there was no difference in their choices to teach an agent or take on a challenge by increasing or decreasing performance conditions. One could interpret this null result as evidence that children use only their final score, rather than their rate of performance, to calibrate future resources. However, many other factors might have affected children's performance here. The procedure likely required preschoolers to represent fairly small differences in relatively large numbers (i.e., the difference between 11, 9, 8, and 7 balls in the Decreasing condition) without explicit counting or labeling. This process relies on children's approximate number system, which develops with age (Halberda & Feigenson, 2008). Thus, children in the Decreasing condition may have not noticed that they got worse over time, but rather computed that they continued to get a roughly high number of balls in on each trial, which might be motivating. Additionally, while the dependent measure in Exps. 1a-c may index an overall feeling of confidence, it does not require children to reason specifically about their future performance. That is, when choosing an easy or hard level for the last trial, an exact goal (you need to get X number of balls in) was not stated. If children instead knew they had to get nine balls in on the final trial, and the paradigm allowed children to track their performance more easily, then children might have been more confident that they could obtain this goal in the Increasing condition than the Decreasing condition.

To better test whether children are sensitive to their slope of performance, and not just their end point, we created a new paradigm that 1) did not rely on children's approximate number system and 2) had a dependent measure that required children to reason about their ability to reach a goal. We also used a constant condition where children's performance did not change over time instead of a decreasing condition, since in the real world, children are more likely to fail to improve than to actually get worse over repeated trials. Finally, we extended our age range to include six-year-olds as older children might be more capable of integrating their past performance into their future choices.

Experiment 2

Methods

Participants & Materials Exp. 2 was pre-registered (https://osf.io/vmu6e). Participants were recruited at an urban children's museum and tested individually in quiet testing rooms off of the museum floor. Seventy 4-6-year-old children were recruited for the study, but only 66 were included in the data analysis (mean age: 63.8 months; range: 48 - 83 months) due to the pre-registered exclusion criteria of missing video recording (n = 2) or parental interference (n = 2). Children were randomly assigned to one of two conditions: Increasing (16 F) or Constant (22 F); age was matched across conditions (age: b = -1.24, 95% CI [-6.23, 3.87]). We preregistered running 33 subjects/ condition from a power analysis anticipating a medium-large effect (V = .4) and a power of 0.9.

Children in both conditions were presented with two visually similar wooden trees with nests on top. One was tall (44.5 cm) and one was small (29.2 cm; note that the trees were never referred to by height to the child, but for clarity, we will refer to the trees as tall and small throughout). Children were tasked with bringing an egg (a metal ball, 2 cm diameter) back up to the nest by putting it on a platform attached to a string on a pulley and pulling the string carefully to raise up the platform. Unbeknownst to the child, the egg stayed on the platform via a hidden magnet in the platform that the experimenter could turn on and off via a remote control. While children pulled up the platform, the ball wobbled, making it appear as a conceivably difficult balancing task. Importantly, this task was novel, minimizing the influence of prior expectations on performance. It was also intuitive and easy for children to track their progress since the experimenter placed marks where the ball fell off on the trunk. (A similar task was used by Stipek et al., 1984 to successfully manipulate performance in four-year-olds.)

Procedure Children in both conditions were introduced to the two trees and told that an egg had fallen out of its nest and that their goal was to put the egg back in the nest. The experimenter told children that they would receive two stickers for getting the egg to the nest in the tall tree and one sticker for getting the egg to the nest in the small tree. The experimenter said that they would start by playing with the tall tree. In both conditions, children played with the tall tree for four consecutive trials. In the Increasing condition, the experimenter surreptitiously made the ball fall off the platform at 8", 10", 12", 14" sequentially during each respective trial. In the Constant condition, the ball fell off around 14" on each trial. After the ball fell off on each trial, the experimenter said, “Oops it fell off! But good job!”, then marked children’s progress by placing a marker with the trial number on the Velcro trunk at the height where the ball fell off (see Figure 2). Then the experimenter prompted the child to try again.

After the fourth trial, the experimenter reviewed the child’s performance by pointing to how high they got the egg on each trial. Finally, the experimenter asked the critical question, “You’re so close to getting to the top of this tree! Do you want to try this tree again (pointing to tall tree) or try the other tree (pointing to small tree)?” After the child made a decision, the experimenter ensured success on whatever tree they chose. The experimenter also asked two follow-up questions – 1. “Why did you choose that tree?” (to make sure they didn’t think the toy was rigged or broken in the constant condition) and 2. “Did you get better at the game over time or stay the same?” (to see if children could explicitly report their progress.) This question wasn’t asked for six children.

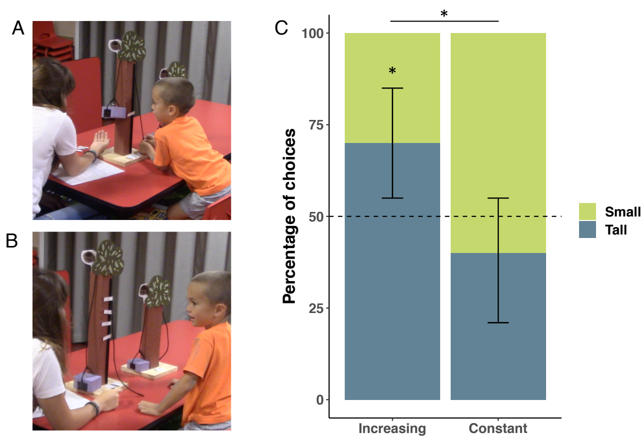

Figure 2. Procedure and results from Exp. 2. A. A child in the Increasing condition on his second trial. B. A child faced with a choice between staying on the tall tree (with markings from his progress in the increasing condition) or switching to the small tree on the right. C. Results from Exp. 2.

Results & Discussion

Children in the Increasing condition chose to continue playing with the tall tree more than children in the Constant condition (χ^2 (1, n = 66) =4.95, p = .03, V = .3; see Figure 2). The percentage of children who chose to play with the tall tree in the Increasing condition was significantly above chance (binomial test against chance (50%): 70% tall, CI = [55%, 85%], p = .04). Children’s choices in the Constant condition did not differ from chance (40% tall, CI = [21%, 55%], p = .30). In a logistic regression predicting tree choice with condition and age, both condition and age were significant (age: b = .07, p =.02; condition: b = 1.32, p =.02), revealing that children in the increasing condition and older children were more likely to choose the tall tree. There was no age by condition interaction on tree choice. The paradigm was convincing to children: None of the children said they chose to switch to the small tree because the tall tree was broken or rigged. Additionally, children in the Increasing condition were more likely to say they got better at the game over time than children in the Constant condition (χ^2 (1, n = 59) = 9.4, p = .002, V = .43; missing data from 7 participants).

Exp. 2 shows that children are sensitive to their rate of improvement, not just their final performance: Children who got closer to reaching a goal over time were more likely to stick with a challenging task than children who were consistently close to reaching a goal over time, even when distance to the goal on the final trial was matched.

General Discussion

Across four experiments, we found that four to six-year-olds are sensitive to their rate of past performance on a task and use this information to allocate future resources. In Exp. 1a and 1b, children who improved at a game over time displayed more confidence (indexed by a combination of being likely to take on a challenging version of the game and teach others about the game) than children who got worse at the game over time. Importantly, children attended to their rate of improvement without any cueing – their trajectory of learning was never explicitly labeled. Furthermore, children weighted their rate and direction of change more than just their total score (Exp. 1a) or final score (Exp. 2) when allocating future effort. Taken together, these findings show that preschool age children attend to their rate of past performance and use this information to guide future efforts.

Past studies have found that young children do possess metacognitive awareness of their past performance on single trials of a task (Goupil et al., 2016; Metcalfe & Finn, 2013; Schneider, 1998), but fail to incorporate this information when planning future actions, leading to over-confident predictions and inefficient time management when learning. Here, we find that given four consecutive data points of progress, children are not “wishful thinkers” or irrational optimists. In line with work on children’s sophisticated rational learning (Schulz, 2012; Tenenbaum et al., 2011), we find that children are able to not only monitor their performance over time, but also use this information to efficiently guide future actions. When children are not improving at a task, they choose to play it again at an easier level and learn more. When children are getting better at a game, they choose to play it again at a more challenging level and teach others about the game. Young children’s optimism after one data point may be advantageous for young learners who often face failure. However, an optimal learner should also be able to effectively allocate resources when evidence suggests that their efforts won’t pay off. Repeated data may help children refine their estimate of their abilities, allowing them to better calibrate their future actions.

The nature of the task may also influence children’s ability to track their progress and calibrate future actions. Here, we purposely used novel tasks that were appropriate for preschoolers and visual aids to lower memory demands and help children track their own progress. These features may be vital to children’s ability to monitor and act on their past performance. Children’s sensitivity to task structure is exemplified in the discrepancy of results between Exp. 1c and 2 – when tracking their own performance required skills arguably beyond young children's abilities and the decision about what to do next didn't have a clearly defined goal, children did not make decisions based on changes in their performance.

When deciding whether to take on a challenge, one must consider both their chance of success and the potential payoff. By manipulating children's performance slopes in Experiments 1 and 2, we likely influenced their estimates of success. However, the potential for rewards differed by Experiment: In Experiment 1, there was no explicit reward for reaching either goal, but in Experiment 2, children could receive a larger reward (two stickers) for completing the more challenging task, which may have shifted their overall preference to stick with a challenge. Yet, in Experiment 2 children's challenge preferences still differed by condition, indicating that children weight not just their potential rewards, but also their chance of getting them. In ongoing work, we are running a version of Experiment 2 with matched rewards to more fully test how children integrate rates of performance with rewards.

Future work would benefit from computational models that can precisely map out how other relevant features are integrated with information about rate (Son & Sethi, 2006). For example, this work controlled for total score (Exp. 1b) and final score (Exp. 1c & Exp. 2), but presumably these features are also important to effort allocation. Furthermore, computational and behavioral work could probe individual differences in response to one's learning trajectory. Some children in Exp. 2 responded negatively to the increasing condition, saying that they kept "failing" on each trial because the ball fell off, even though they were getting closer to the end goal. Individual differences in children's reaction to failure, as well as their interest in the task, the long term pay off of mastering the task, and who is watching, may all influence children's interpretation of their learning curve.

This work has a number of limitations. In Experiments 1a-c, we combined the basket choice and puppet choice outcome measures into one competence score. These two measures index different aspects of competence (challenge and social interaction preference), which we thought might collectively vary by condition. However, the condition effect on these two measures was inconsistent across Experiments 1a-c, suggesting that combining these measures may not have been appropriate. Subsequently, we decided to focus on just challenge preference in Experiment 2. The children in Experiment 2 were also on average a year older than those in Experiment 1, raising the question of whether children's sensitivity to performance over time changes with age. Indeed, in Experiment 2 we found that older children were more likely to take on a challenge than younger children across conditions. Ongoing work is exploring how differences in children's ability to track and predict performance rate by age might contribute to age related changes in challenge preference. Finally, there is a chance that children in the Constant condition in Experiment 2 were switching tasks due to boredom, rather than a richer interpretation of their past performance. Future work should probe children's challenge preference using a stochastic Constant condition, with performance varying slightly, to test if boredom is potentially driving this condition effect.

The ability to track learning over time is an essential element of efficient self-directed learning. Our work shows that young children possess this ability, which may lend to their remarkable capacity to learn so much with so little experience. Academic settings present a situation where this form of learning often has real-world consequences. This work suggests that young children may benefit from intentional scaffolding that highlights their progress over time, instead of just feedback on performance on a given day.

Acknowledgments

This work was supported by a MindCORE postdoctoral research fellowship from the University of Pennsylvania to J.A.L and the Jacobs Foundation Early Career Award to A.P.M. We thank the Please Touch Museum and the Boston Children's Museum and participating families. We also thank Fatima Gunter-Rahman, Skyler Cordrey, and Greer Bizzell-Hatcher for their help with data collection. We especially thank Rachel Magid for her help designing and analyzing Experiments 1a-c.

References

Bonawitz, E. B., van Schijndel, T. J. P., Friel, D., & Schulz, L. (2012). Children balance theories and evidence in exploration, explanation, and learning. *Cognitive Psychology*, *64*(4), 215–234. https://doi.org/10.1016/j.cogpsych.2011.12.002

Dunlosky, J., & Nelson, T. O. (1992). Importance of the kind of cue for judgments of learning (JOL) and the delayed-JOL effect. *Memory & Cognition*, *20*(4), 374–380. https://doi.org/10.3758/BF03210921

Flavell, J. H., Friedrichs, A. G., & Hoyt, J. D. (1970). Developmental changes in memorization processes. *Cognitive Psychology*, *1*(4), 324–340. https://doi.org/10.1016/0010-0285(70)90019-8

Goupil, L., Romand-Monnier, M., & Kouider, S. (2016). Infants ask for help when they know they don't know. *Proceedings of the National Academy of Sciences*, *113*(13), 3492–3496. https://doi.org/10.1073/pnas.1515129113

Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the "number sense": The approximate number system in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental Psychology*, *44*(5), 1457–1465. https://doi.org/10.1037/a0012682

Leonard, J. A., Lee, Y., & Schulz, L. E. (2017). Infants make more attempts to achieve a goal when they see adults persist. *Science*, *357*(6357), 1290–1294. https://doi.org/10.1126/science.aan2317

Lucca, K., Horton, R., & Sommerville, J. A. (2020). Infants rationally decide when and how to deploy effort. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-019-0814-0

Metcalfe, J., & Finn, B. (2013). Metacognition and control of study choice in children. *Metacognition and Learning*, *8*(1), 19–46. https://doi.org/10.1007/s11409-013-9094-7

Metcalfe, J., & Kornell, N. (2005). A Region of Proximal Learning model of study time allocation. *Journal of Memory and Language*, *52*(4), 463–477. https://doi.org/10.1016/j.jml.2004.12.001

Nelson, T. O., & Leonesio, J. R. (1988). *Allocation of self-paced study time and the "labor-in-vain effect."* Journal of Experimental Psychology: Learning, Memory, and Cognition. https://doi.org/10.1037/0278-7393.14.4.676

Parsons, J. E., & Ruble, D. N. (1977). The Development of Achievement-Related Expectancies. *Child Development*, *48*(3), 1075–1079. JSTOR. https://doi.org/10.2307/1128364

Rhodes, M., & Brickman, D. (2008). Preschoolers' Responses to Social Comparisons Involving Relative Failure. *Psychological Science*, *19*(10), 968–972. https://doi.org/10.1111/j.1467-9280.2008.02184.x

Schneider, W. (1998). Performance prediction in young children: Effects of skill, metacognition and wishful thinking. *Developmental Science*, *1*(2), 291–297. https://doi.org/10.1111/1467-7687.00044

Schulz, L. (2012). The origins of inquiry: Inductive inference and exploration in early childhood. *Trends in Cognitive Sciences*, *16*(7), 382–389. https://doi.org/10.1016/j.tics.2012.06.004

Siegel, M. H., Magid, R., Tenenbaum, J. B., & Schulz, L. E. (2014). Black boxes: Hypothesis testing via indirect perceptual evidence. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *36*(36), 7.

Smiley, P. A., & Dweck, C. S. (1994). Individual differences in achievement goals among young children. *Child Development*, *65*(6), 1723–1743. https://doi.org/10.1111/j.1467-8624.1994.tb00845.x

Son, L. K., & Sethi, R. (2006). Metacognitive Control and Optimal Learning. *Cognitive Science*, *30*(4), 759–774. https://doi.org/10.1207/s15516709cog0000_74

Stahl, A. E., & Feigenson, L. (2015). Observing the unexpected enhances infants' learning and exploration. *Science*, *348*(6230), 91–94. https://doi.org/10.1126/science.aaa3799

Stipek, D. J., & Hoffman, J. M. (1980). Development of Children's Performance-Related Judgments. *Child Development*, *51*(3), 912–914. JSTOR. https://doi.org/10.2307/1129485

Stipek, D. J., Roberts, T. A., & Sanborn, M. E. (1984). Preschool-Age Children's Performance Expectations for Themselves and Another Child as a Function of the Incentive Value of Success and the Salience of Past Performance. *Child Development*, *55*(6), 1983–1989. JSTOR. https://doi.org/10.2307/1129773

Stipek, D., Recchia, S., McClintic, S., & Lewis, M. (1992). Self-Evaluation in Young Children. *Monographs of the Society for Research in Child Development*, *57*(1), i–95. JSTOR. https://doi.org/10.2307/1166190

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to Grow a Mind: Statistics, Structure, and Abstraction. *Science*, *331*(6022), 1279–1285. https://doi.org/10.1126/science.1192788

Young, A., & Fry, J. D. (2008). Metacognitive awareness and academic achievement in college students. *Journal of the Scholarship of Teaching and Learning*, *8*(2), 10.