Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Two-year-olds can reason about the temporal structure of their performance

Permalink

https://escholarship.org/uc/item/1vb912rp

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)

Authors

Zhu, Peter
Aranda, Veronica
Keene, Grace E
et al.

Publication Date

2024

Two-year-olds can reason about the temporal structure of their performance

Peter Zhu, Veronica Aranda, Grace Keene, & Hyowon Gweon
{pgzhu, varanda, gkeene, hyo}@stanford.edu
Department of Psychology, Stanford University
Stanford, CA, USA

Abstract

When learners improve, the temporal change in performance carries information about progress; we know we "got the hang of it" after succeeding on a task we used to fail at. Building on prior work investigating older children's ability to track their performance over time, here we ask whether two-year-olds can reason about the temporal pattern of their performance outcomes. Children in the Improvement condition experienced 3 failures followed by 3 successes (FFFSSS) whereas children in the Stochastic condition experienced the same number of failures and successes in a seemingly random sequence (SSFFSF). When asked which toy they wanted to show their parent, children were more likely to select the Test Toy over a Control Toy when the temporal sequence of their performance suggested improvement than when it appeared to be random. By reasoning about their own performance over time, even young children can make informed choices about their future actions.

Keywords: Cognitive development; Performance; Ability; Learning progress

Introduction

We often hear the phrase "Practice makes perfect", an old adage that emphasizes the value of persistence and repetition in mastering a task. But how do we know whether we're actually making progress? The ability to reason about our own progress allows us to predict our future performance, decide whether to keep trying or give up, and even choose which goals to pursue next. This is especially important for young children, who constantly face novel tasks and need to master new skills. Despite its importance, however, much remains unknown about whether young children can represent their progress and use it to infer what they can (or cannot) do, and how this ability develops in early childhood.

As adults, we intuitively understand that progress in learning is reflected in how our performance outcomes change over time. Imagine you are learning how to ski; over time, you may notice changes in how fast you can go, how smooth your turns are, or how fatigued you are after a long day at the slopes. Along with these continuous or graded signals, you might also notice changes in how often you succeed or fail. If you transition from falling several times to consistently making it down without a single fall, this temporal distribution of your past failures and successes would suggest that you "got better" at skiing over time, and it may even help you decide when to move onto a more challenging slope. While the nature and the kinds of available signals might depend on the task and the context, the temporal structure in these signals often carries useful information about progress.

A body of prior work in psychology and education suggests that people can monitor and judge their past performance and learning. For instance, adults can use such information to guide predictions of their performance (e.g., test performance, Maki & Berry, 1984) and allocate effort on future tasks, such as appropriately apportioning study time to prioritize items they are more uncertain about (e.g., Masur et al., 1973; Son & Metcalfe, 2000). The process of judging one's own knowledge and learning can also have a positive influence on achievement and self-efficacy (e.g., Young & Fry, 2008; Dunlosky & Nelson, 1992; Panadero et al., 2017; Gürel et al., 2020). Collectively, this literature demonstrates that by adulthood, humans develop a nuanced understanding of how their learning progresses over time, and can use it to predict their future performance or decide what to do next.

In the field of cognitive development, young children's ability to reason about their own learning and knowledge has largely been explored in the context of meta-cognitive development. For instance, classic research has examined how children monitor their ability to recall items from memory (Flavell et al., 1970) or assess their comprehension of task instructions (Markman, 1977). More recent work has looked at how children reflect on their learning process; for instance, around 7 to 8 years of age, children begin to describe learning as a process of acquiring knowledge or skills (Sobel & Letourneau, 2015) and expect that they will make graded improvement at a task over successive attempts (Zhang et al., 2023). These studies, however, primarily focused on relatively older children already in formal schooling, and often required the ability to verbalize their own uncertainty, reflect upon their learning, or reason about outcomes that are relatively fine-grained and task-specific. While these measures can probe the learners' metacognitive awareness of their own learning, they also impose significant cognitive demands such as remembering and reflecting on past experiences, or even require a priori knowledge about which signals carry information about their progress on specific tasks.

One recent study has directly explored whether 4-to 6-year-olds are sensitive to changes in their past performance over time (Leonard et al., 2023). In this study, children tackled a challenging task that involved getting an "egg" into a nest at the top of a wooden tree; they had to use a pulley device to raise a platform containing the toy egg to a specific height (17.5") on a tree-shaped device. While the task was rigged

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society*.

such that all children failed repeatedly at the task, those who experienced increasingly better outcomes (i.e., closer to the target height, from 8" to 10", 12", and 14") were more likely to attempt a harder task (i.e., a tall tree over a short one) than children who consistently failed at 14". In other words, children were able to predict their future performance by extrapolating from the temporal change in their past performance. Unlike other studies that relied on explicit self-reflection, this study leveraged children's future task choice as a way to assess whether preschool-aged children can reason about their past performance; the underlying assumption is that children would be more likely to choose a harder task insofar as they expect to perform better in the future.

One notable aspect of the task used in Leonard et al. (2023) is that the evidence of improvement—the pattern of change in performance outcomes over time—involved a linear, gradual decrease between the outcome and the target. Indeed, the concrete, physical nature of this signal may have been salient and easy enough to understand even for preschool-aged children, allowing researchers to identify competence in this age group. However, in order to use this information, children also need a task-specific understanding of the overall goal and the mechanics of the device (i.e., lifting an egg to the top of a tree) and to interpret the graded signal as an indicator of their progress. While even infants can represent numerical quantities (e.g., Feigenson et al., 2004) and relative distance (e.g., de Hevia & Spelke, 2010), grasping the meaning of these quantities with respect to a goal might be challenging for infants and toddlers, especially given their limited linguistic ability to understand verbal task instructions.

Learners' past performance outcomes, however, offer another—and arguably simpler—signal that even younger children might understand: the temporal distribution of one's past failures and successes. If you consistently succeed at a task you used to fail at, you may infer that you are progressing in your learning, or even that you "got the hang of it". Indeed, infants and toddlers routinely engage in manual exploration of causal toys, and in the face of failure, make repeated attempts to succeed. Such behaviors have been leveraged as dependent measures in past research to assess inductive inference (Gweon et al., 2010) and persistence (Leonard et al., 2017) in preverbal infants. However, whether and when children can use the temporal distribution of these outcomes as a signal of progress remains an open question.

Some existing work lends support to the possibility that this ability emerges quite early in life. First, a large body of research shows that the ability to extract statistical patterns from temporal signals (e.g., speech) is already present in the first year of life (see Gómez & Gerken, 2000; Saffran & Kirkham, 2018, for reviews); thus, by two years of age, children may well be sensitive to the temporal patterns in other signals, such as their own performance outcomes. Second, by 20 months of age, toddlers ask for help when they don't know, demonstrating an early-emerging sense of their own uncertainty (Goupil et al., 2016). Finally, beyond distinguishing cases where another agent succeeds vs. fails to reach a goal (e.g., Hamlin et al., 2007, 2008; Brandone & Wellman, 2009), by 18 months of age, infants respond differently to their *own* failures and successes (e.g., saying "uh-oh", Gopnik, 1982; Gopnik & Meltzoff, 1986), and even leverage the covariance structure in others' failures and successes to guide their future actions (Gweon & Schulz, 2011).

Importantly, however, these studies did not directly manipulate the temporal distribution of participants' own failures and successes. A large body of work has also documented children's overconfidence and optimism bias regarding their own performance that persists throughout early childhood (e.g., Schneider, 1998; Hennefield & Markson, 2022; Lipko et al., 2009), which may hinder or mask children's ability to reason about their own performance until late preschool years. In light of these possibilities, here we investigate whether two-year-olds are sensitive to the temporal pattern of their past performance outcomes and can use it to guide their future behaviors.

To this end, we designed a study where children had experience with two toys: a Control Toy and a Test Toy. While all children succeeded consistently on the Control Toy, they experienced both failures and successes with the Test Toy. What differed across conditions was the sequence of their failures and successes on the Test Toy: children either consistently failed first then consistently succeeded (Improvement condition) or experienced a seemingly random sequence (Stochastic condition). In order to assess whether toddlers are sensitive to such patterns, participants were encouraged to choose a toy to show to their parent. We predicted that children would be more likely to choose the Test Toy in the Improvement condition than in the Stochastic condition. We first report a preliminary (pilot) study to verify a key assumption in our dependent measure (Experiment 1), followed by a preregistered study (Experiment 2).

Experiment 1: Preliminary Study

In this preliminary study, we collected a pilot sample in order to understand how toddlers would respond to our task. Our key prediction was that children would be more likely to choose the Test Toy to show their parent (over the Control Toy) when the temporal sequence of their past performance indicated improvement than when it was seemingly random. This prediction critically hinges on an assumption: Children would be motivated to demonstrate a toy they *think* they can activate over a toy they are less certain about. Prior work suggests that two-year-olds modulate their engagement with a toy based on others' attention and expressed preferences (e.g., Botto & Rochat, 2018); by early preschool years, children also show a clear desire to demonstrate their success to others (Asaba & Gweon, 2022). Thus, we expected that children might be more reluctant to choose the Test Toy over the Control Toy when their experience with the Test Toy varied randomly than when it signaled progress.

However, it is possible that children are also motivated to

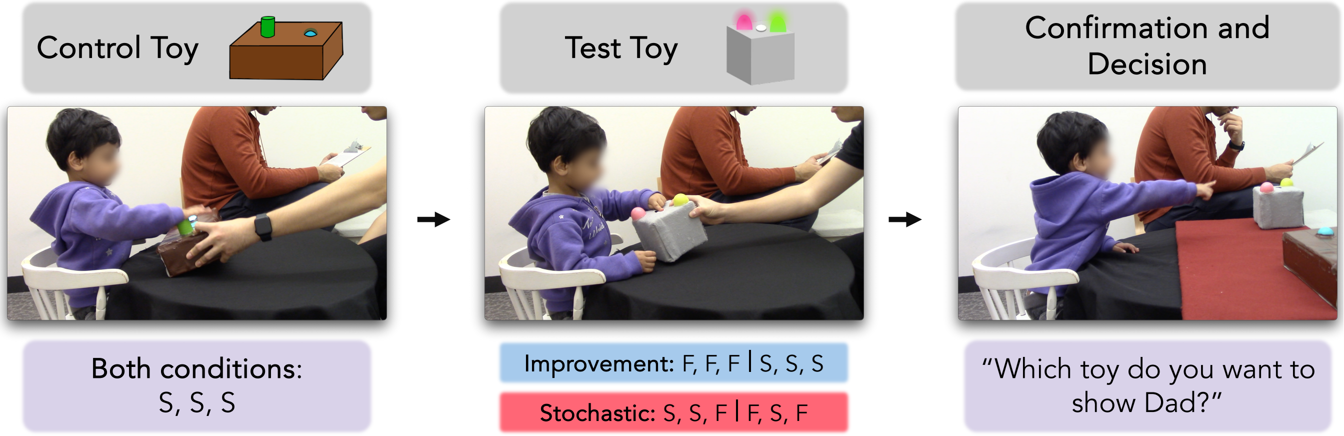

Figure 1: Study procedure. The child's caregiver sat next to the child but was turned away, wearing a set of headphones. Control Toy Phase: All children successfully activated the Control Toy three times. Test Toy Phase: Children attempted the Test Toy six times (3 failures, 3 successes), but the temporal structure of the outcomes differed across conditions. Confirmation and Decision Phase: Children successfully activated each toy once, and selected a toy they want to show their caregiver.

explore a toy they are *less* certain about, or one that produces a pattern that elicits surprise or curiosity. Indeed, young children tend to explore more when they observe inconsistent events or ambiguous patterns of data (e.g., Stahl & Feigenson, 2015; Legare, 2012; Schulz & Bonawitz, 2007). Thus, children's desire to explore the toy that resulted in a seemingly random sequence of outcomes might either cancel out or even override their desire to generate a successful outcome.

A difference between conditions, regardless of its directionality, would be consistent with the hypothesis that children are sensitive to the temporal pattern of their past performance. However, these competing motivations make it challenging to interpret the meaning of their choice. In order to justify the aforementioned assumption in our dependent measure, we first ran a pilot sample to identify the broader trend in the data, followed by a preregistered study. While we aimed our final study to focus on two-year-olds, given that clear evidence of children's motivation to demonstrate their success comes from studies of 3- to 5-year-olds (Asaba & Gweon, 2022), we recruited both 2- and 3-year-olds for this pilot study.

Methods

Participants Sixteen (n = 8/condition) 2- and 3-year-old children were tested at a local children's museum ($M(SD)_{age}$ = 2.77 (0.62) yrs, Range: 1.80 - 3.88 yrs). An additional 6 children were excluded due to not finishing the study (n = 3) or a faulty toy that did not activate as planned (n = 3).

Stimuli Two toys—a Control Toy and a Test Toy—were constructed using foam boards, felt cloth, foam cylinders, LED lights, and electronic buttons (see Figure 1). The Control Toy was a brown rectangular box with a green foam cylinder and a blue LED light on the top surface. The Test Toy was a gray box with a white button and two LED lights (green and pink) on the top surface. When activated, the LED lights on the toys lit up; although pressing the green foam cylinder (Control Toy) or pressing the white button (Test Toy) appeared to activate the toys, the lights were surreptitiously operated by the experimenter using a remote control hidden under the table. We deliberately designed the Test Toy to be more perceptually appealing than the Control Toy (e.g., two lights rather than one) such that activating it would be relatively more rewarding than activating the Control Toy. Additionally, for the toy choice measure, a foam board covered in red felt was used to simultaneously present the two toys.

Procedure All children were tested in a quiet room at the children's museum. Children were seated in a high chair pulled up against a table, with the experimenter sitting directly across from them and their caregiver seated perpendicular to them. The experimenter instructed the caregiver to turn to the side and put on a set of headphones. The experimenter made it clear to the child that their caregiver will not be attending to them during the procedure by saying: "[Mom/Dad] is going to listen to music, so say bye!". The caregiver then turned their body away from the table, such that they appeared to not be observing the study. The procedure unfolded in two phases.

1. Control Toy Phase: The experimenter brought out the two toys for the main experiment: a Control Toy (brown) and a Test Toy (gray; Left/Right position counterbalanced). The experimenter offered to play with the Control Toy first, moving it to the center and removing the Test Toy from the table.

The experimenter demonstrated the toy's function twice by placing their hand on the green cylinder, which appeared to activate a blue light (the light was surreptitiously controlled by the experimenter). The experimenter then offered the toy to the child; the child successfully activated the toy three times (indicated as SSS in Figure 1). After each success, the experimenter displayed a positive reaction and encouraged the child to try again ("Ooh! How about you try again?").

2. Test Toy Phase: After the child activated the Control Toy for the third time, the experimenter removed it from the table and brought out the Test Toy. The experimenter demonstrated the Test Toy in the same manner as the Control Toy, activating it twice by pressing the white button on top which appeared to activate both the pink and green lights. The child was then given six opportunities to try the toy. Critically, while all children failed three times and succeeded three times, the temporal structure of their performance differed across conditions: Children in the Improvement Condition failed three times and then succeeded (FFF|SSS) whereas children in the Stochastic condition experienced what appeared to be a random sequence (SSF|FSF). The experimenter displayed a positive reaction ("Ooh!") to the child's success as in the Control Toy phase, and a puzzled reaction ("Hmm...") to the failure.

The '|' in the sequence indicates a brief demonstration by the experimenter; after the child's third attempt on the Test Toy (i.e., after FFF in the Improvement condition, or SSF in the Stochastic condition), the experimenter said that the child would have to press 'really hard' to activate the toy, and demonstrated its effect once more before giving it back to the child to try three more times (i.e., SSS in the Improvement condition, or FSF in the Stochastic condition). The purpose of this demonstration was twofold. For children in the Improvement condition, it could provide a plausible reason for their sudden success (and past failure); for children in the Stochastic condition, continuing to fail even after the experimenter's demonstration, despite their effort to press harder, would likely heighten the uncertainty about their ability to activate the toy. Note, however, that understanding the verbal instruction was not critical to the study; it was included to maximize the effect of the temporal sequence by increasing their uncertainty (in the Stochastic condition) or their expectation of success (in the Improvement condition).

3. Confirmation & Choice Phase: After the child's sixth attempt on the Test Toy, the experimenter presented both the Control Toy and the Test Toy on a tray and asked the child to activate each toy again. The relative position of the toys were counterbalanced, and the experimenter always asked the child to activate the toy on the left first. All children successfully activated both toys, such that the recency of their final success with the toys was matched across conditions.

Finally, the experimenter said, "We're going to play with [Mom/Dad] now! Do you want to show [Mom/Dad] how to make one of these toys go? You can show [Mom/Dad] how to make *only one* of these toys go. Which toy do you want to show [her/him]?"). The key dependent measure was which toy children selected to show their caregiver, indicated by pointing or reaching at either the Control or the Test Toy.

Results and Discussion

While all children in the Improvement condition picked the Test Toy over the Control Toy (8 of 8, 100%), only half of the children in the Stochastic condition (4 of 8, 50%) chose the Test Toy. This difference did not reach statistical significance (p = .077; Fisher's Exact Test), and we were likely underpowered given the small sample size. Nonetheless, these preliminary results alleviate the concern that children's curiosity about their performance on the Test Toy in the Stochastic condition (i.e., seemingly random outcomes) would drive children's choices; instead, our choice task (i.e., choosing a toy to show their caregiver) likely led children to choose a toy they think they can reliably activate.

The study was designed to be robust against a few obvious low-level explanations. First, the number of successes and failures that children experienced on both toys was matched across conditions (three successes for the Control Toy, and three failures and three successes for the Test Toy). Thus, the results cannot be attributed to the relative frequency of failures and successes. Furthermore, just before the final choice, all children successfully activated both toys in counterbalanced order (Confirmation & Choice Phase), such that the recency of their final success cannot explain children's choice.

In this pilot sample, all children in the Improvement condition chose the Test Toy over the Control Toy. It is important to note that there is no a priori reason to expect such a strong preference. In fact, children in both conditions likely had good reasons to choose the Control Toy; although we designed it to be perceived as somewhat less rewarding than the Test Toy (i.e., it had one light rather than two), it still produced an enjoyable effect, and children never failed to activate it. Compared to the Test Toy that resulted in three failures in both conditions, children may have had relatively high confidence in their ability to activate the Control Toy again. Thus, our key hypothesis for the main study focuses on a condition difference rather than whether children's toy choice deviates from chance (50%).

Taken together, these pilot results suggested that even two-year-old children can understand our task and are motivated to choose a toy that they think they can successfully activate when asked to show a toy to their parent. In Experiment 2, we sought to provide more robust evidence for our hypothesis by replicating these findings with a larger, preregistered sample.

Experiment 2: Preregistered Study

Given the pilot results and the absence of a clear effect of age, we limited our age range to two-year-olds. Also, given the relatively robust trend in Expt. 1, we preregistered a sequential Bayes Factor (BF) analysis (e.g., Mani et al., 2021) using the `BFpack` package in R (Mulder et al., 2019).[1] Following the preregistered plan, after an initial sample of $n = 40$, the

---

[1] We used `BFpack`'s approach for computing the BF for generalized linear models, which calculates the adjusted fractional BF using Gaussian approximations (Gu et al., 2018). Planned sample size, analyses, and exclusion criteria were preregistered here. Data and analysis scripts can be found here.

BF was evaluated after every four participants, with a plan to stop data collection when one of the following conditions were met: strong evidence in favor of the hypothesis of a condition effect ($BF_{10} > 10$), weak evidence in favor of the null hypothesis ($BF_{10} < 1/3$), or reaching a pre-planned *n* = 72. Thus, this approach allowed us to take a principled, data-driven approach for optionally stopping data collection before reaching the final sample size in the presence of strong evidence for either the predicted or null hypothesis.

Methods

Participants Our sequential Bayes Factor approach led to a final *n* of 60 two-year-old children (*n* = 30/condition), who were recruited from a local children's museum or a nursery school on campus ($M(SD)_{age}$ = 2.52 (.31) yrs, Range: 2.01 - 3.05 yrs). We excluded an additional 23 participants based on the following pre-registered exclusion criteria: experimenter error (*n* = 7), failing to finish the study (e.g., fussing out, *n* = 5), tech error (e.g., faulty toy, *n* = 5), parental interference (*n* = 2), or failing to make a choice at the Decision Phase (e.g., not selecting a toy when prompted, *n* = 4).

Stimuli The stimuli were identical to Experiment 1, except that we replaced the high chair with a small arm chair in order to help children feel more comfortable and minimize the number of children who would refuse to sit in the chair.

Procedure The procedure for Experiment 2 was largely identical to Experiment 1. However, given our focus on two-year-olds, we implemented a few minor changes to maximize the possibility that our participants could understand the task.

First, the procedure started with a brief warm-up phase. Just after their caregiver turned away, children were presented with two rubber toys (a red starfish and an orange fish) on the red tray. The experimenter then told children that they could play with "only one of these toys" and instructed them to point at the toy that they would like to play with (the wording was identical to the test question in the Decision Phase). Once children selected their toy, the experimenter gave them the toy they chose (removing the tray and the unchosen toy) and played with the child for approximately 5 seconds. Thus, the warm-up phase allowed children to practice pointing as a way to indicate their choice, and established an expectation that they would only receive the toy that they had pointed to.

Second, during the Decision Phase, we added various behavioral cues that would help children understand that they were choosing a toy to show to their caregiver. For instance, the experimenter explicitly signaled towards the child's caregiver while telling the child they could show one of the toys to their caregiver, and similar to the choice practice during the warm-up phase, the experimenter simultaneously pointed at the Test Toy and the Control Toy to elicit children's choice.

Results

The results were consistent with the pattern observed in Experiment 1: Children in the Improvement condition were more likely to pick the Test Toy over the Control Toy (20 of 30; 66%) than children in the Stochastic condition (9 of 30; 30%). Our planned logistic regression (assessing the data in support of the hypothesis of a condition effect over the null hypothesis of no difference) yielded a Bayes Factor (BF_{10}) of 12.02, which we interpret as strong evidence in favor of the predicted hypothesis. As an exploratory analysis, we conducted a Fisher's Exact Test in order to test the effect of condition on choice. This analysis revealed a significant effect of condition ($OR = 4.54$, 95% CI: [1.39, 16.04], $p = .009$).

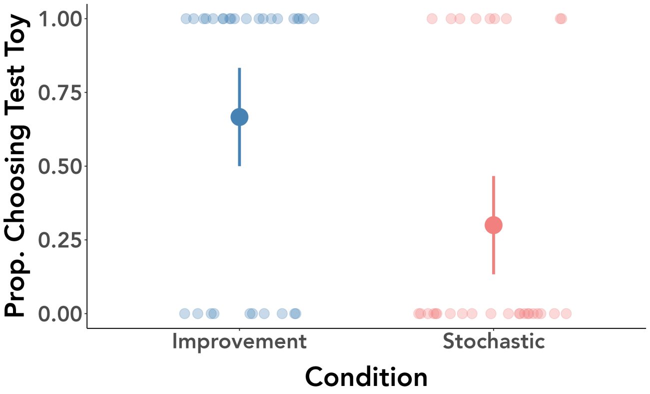

Figure 2: Experiment 2 results: Proportion of children who chose the Test Toy in each condition. Dots represent individual participants; error bars are 95% bootstrapped CIs.

General Discussion

The current study examined two-year-olds' ability to reason about the temporal structure of their failures and successes. To measure this, we looked at which one of two toys they chose to show their caregiver. We found that children's choice of toys was modulated by the temporal structure of their past failures and successes; children were more likely to choose the Test Toy (over the Control Toy which always resulted in success) when their experience with the Test Toy indicated progress over time (i.e., consistent failure followed by consistent success; Improvement Condition) than when their experience was seemingly random (Stochastic condition). Importantly, this difference across conditions cannot be explained by the sheer frequency of failures and successes or the recency of success; the number of failures and successes were matched across conditions, and all children successfully activated both toys immediately before the final choice.

These findings build on and extend existing findings on children's early-emerging ability to leverage the statistical regularities in their environment. While prior work has focused on exogenous signals such as sequences of speech syllables (e.g., Saffran & Kirkham, 2018), visual displays (e.g., Bulf et al., 2011), or covariance structure in successes and failures across agents and causal mechanisms (Gweon & Schulz, 2011), the current work focuses solely on the outcomes of children's own failures and successes. Our results suggest that by age 2, children readily distinguish different temporal structures in their past performance outcomes and use these patterns to guide their future behaviors.

In the current study, we were interested in how children utilize information about their past performance in order to make inferences about their own abilities (i.e., "can I make this toy go?"). However, one alternative explanation for the results is that children's choice simply reflects their response to the toys themselves, rather than their reasoning about their own abilities. That is, children in the Stochastic condition may have been frustrated with the seemingly random activation of the Test Toy and therefore were more likely to *avoid* the Test Toy compared to those in the Improvement condition. While the current results do not completely rule out this possibility, it is rather unlikely given our experimental design. Specifically, the experimenter activated the Test Toy three times (twice during the initial demonstration, once during the re-teaching of the toy) without fail. This was a deliberate design decision to mitigate the concern that children would think the toy is faulty, and provide contextual support for the inference that they themselves (rather than the toy) are the cause of their failures (Gweon & Schulz, 2011). Yet, this alternative account nonetheless raises an interesting direction for future work. While the current study motivates children's choice with a presentational goal (i.e., the desire to appear positively to others) by asking them to choose a toy to show to their caregiver, introducing a different goal could alter the pattern of choice. For instance, if children are asked to choose a toy to keep playing with by themselves, would they show a relatively greater preference for the Test Toy even in the Stochastic condition? Such a preference could reflect a desire to explore a toy that produced variable outcomes, particularly in the Stochastic condition. These results would not only rule out the possibility that children in the Stochastic condition had a general aversion to the Test Toy, but also would suggest that children encode the structure in their outcomes and flexibly adapt their behavior depending on the goal.

These findings complement prior work using graded outcomes (e.g., Leonard et al., 2023) by leveraging a relatively more rudimentary representation of binary outcomes (i.e., failures and successes). While this approach helped us minimize task demands for younger participants, whether toddlers can also leverage other signals in their learning remains an important open question. For instance, research on early motor learning suggests that even infants might be sensitive to internal signals (e.g., proprioception) that have more complex, graded patterns (e.g., Adolph, 2008, 2000). This raises the possibility that toddlers can also harness more graded aspects of their progress (e.g., "how far did I get to my goal?"). On the other hand, it is also plausible that toddlers simply compress their past performance to abstract representations of success or failure (e.g., "did I reach my goal or not?"). Future work might investigate how toddlers harness a variety of learning signals as well as the nature of the representations that underlie such learning about the self.

Relatedly, our findings also raise an interesting question about our representations of success and failure: What are the cognitive processes that underlie our ability to abstract away from our concrete experiences and conceptualize them as "success" or "failure"? In the example of skiing, while any arbitrary criteria could, in principle, be used to demarcate success, there is likely substantial systematicity in what people construe as a reasonable set of criteria. These criteria also vary depending on the task, the learner, and the context. While a complete novice might represent success as safely making it down the beginner slope, as their skill grows, success might mean completing a difficult slope under tight time pressure. Indeed, how learners reason about "success" is often subjective; while there are cases with clear delineations (e.g., activating a toy), other times learners have to define it for themselves. Although prior work has largely investigated children's understanding of failed action in contexts of clear, but incomplete, goal-directed action (e.g., failing to reach an object, Brandone & Wellman, 2009; Hamlin et al., 2008), learning in real-world contexts often involve tasks without obvious, concrete markers of success or failure. In these cases, information from others (e.g., emotional reactions or praise from caregivers or teachers) may play an important role. We look forward to future work that aims to better understand how people conceptualize success and failure and how this ability develops in early childhood.

Importantly, reasoning about one's own performance over time is more than just thinking about the successes and failures, or even paying attention to more subtle cues of improvement; feelings of "getting it" or "being clueless" are often also connected with distinct affective states. Even as adults, we experience this often in our daily lives; progressing towards a learning goal can feel good, while unexplained failures may be frustrating. While some work has begun to explore the cognitive and neural bases of "A-ha!" moments (i.e., the feeling after reaching a solution to a complex problem, Kounios & Beeman, 2009; Danek et al., 2014; Dubey et al., 2021), how this feeling translates to gradual progress over time remains an open question. Utilizing methods such as affective facial coding or physiological measures might help us gain some empirical traction on how such signals might serve as indicators of progress or stagnation in learning.

A key aspect of any self-guided learning or exploration is an understanding of one's own abilities. Questions like "what can I do?" or "what am I good at?" reflect our desire to understand what we can (or cannot) do, when we should seek help, and what challenges we should pursue next. Complementing recent work on children's desire to seek information about the self from others (Zhu et al., 2023), here we show that children are also able to reason about their abilities through self-generated evidence. Even though these children are still too young to succeed on many of the tasks they see others succeed on, they are nonetheless able to reason about their own performance and use it to inform their future actions. These abilities may serve as an important foundation for learning and improvement throughout childhood and beyond.

Acknowledgments

We thank Ellen Aasted and Libby Rouffy for their help with data collection and the Social Learning Lab for helpful comments and feedback. We also thank Bing Nursery School and the Palo Alto Junior Museum and Zoo: thank you to Chiawa Yeh for help recruiting and supervising testing, and to the families and children who participated. This work was funded by a James S. McDonnell Scholar Award to HG.

References

Adolph, K. E. (2000). Specificity of learning: Why infants fall over a veritable cliff. *Psychological Science*, *11*(4), 290–295.

Adolph, K. E. (2008). Learning to move. *Current directions in psychological science*, *17*(3), 213–218.

Asaba, M., & Gweon, H. (2022). Young children infer and manage what others think about them. *Proceedings of the National Academy of Sciences*, *119*(32), e2105642119.

Botto, S. V., & Rochat, P. (2018). Sensitivity to the evaluation of others emerges by 24 months. *Developmental Psychology*, *54*(9), 1723.

Brandone, A. C., & Wellman, H. M. (2009). You can't always get what you want: Infants understand failed goal-directed actions. *Psychological science*, *20*(1), 85–91.

Bulf, H., Johnson, S. P., & Valenza, E. (2011). Visual statistical learning in the newborn infant. *Cognition*, *121*(1), 127–132.

Danek, A. H., Fraps, T., von Müller, A., Grothe, B., & Öllinger, M. (2014). It's a kind of magic—what self-reports can reveal about the phenomenology of insight problem solving. *Frontiers in psychology*, *5*, 1408.

de Hevia, M. D., & Spelke, E. S. (2010). Number-space mapping in human infants. *Psychological science*, *21*(5), 653–660.

Dubey, R., Ho, M. K., Mehta, H., & Griffiths, T. (2021). Aha! moments correspond to meta-cognitive prediction errors.

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*, 374–380.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in cognitive sciences*, *8*(7), 307–314.

Flavell, J. H., Friedrichs, A. G., & Hoyt, J. D. (1970). Developmental changes in memorization processes. *Cognitive psychology*, *1*(4), 324–340.

Gómez, R. L., & Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in cognitive sciences*, *4*(5), 178–186.

Gopnik, A. (1982). Words and plans: Early language and the development of intelligent action. *Journal of Child Language*, *9*(2), 303–318.

Gopnik, A., & Meltzoff, A. N. (1986). Relations between semantic and cognitive development in the one-word stage: The specificity hypothesis. *Child Development*, 1040–1053.

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.

Gu, X., Mulder, J., & Hoijtink, H. (2018). Approximated adjusted fractional bayes factors: A general method for testing informative hypotheses. *British Journal of Mathematical and Statistical Psychology*, *71*(2), 229–261.

Gürel, Ç., Brummelman, E., Sedikides, C., & Overbeek, G. (2020). Better than my past self: Temporal comparison raises children's pride without triggering superiority goals. *Journal of Experimental Psychology: General*, *149*(8), 1554.

Gweon, H., & Schulz, L. (2011). 16-month-olds rationally infer causes of failed actions. *Science*, *332*(6037), 1524–1524.

Gweon, H., Tenenbaum, J. B., & Schulz, L. E. (2010). Infants consider both the sample and the sampling process in inductive generalization. *Proceedings of the National Academy of Sciences*, *107*(20), 9066–9071.

Hamlin, J. K., Hallinan, E. V., & Woodward, A. L. (2008). Do as i do: 7-month-old infants selectively reproduce others' goals. *Developmental science*, *11*(4), 487–494.

Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature*, *450*(7169), 557–559.

Hennefield, L., & Markson, L. (2022). The development of optimistic expectations in young children. *Cognitive Development*, *63*, 101201.

Kounios, J., & Beeman, M. (2009). The aha! moment: The cognitive neuroscience of insight. *Current directions in psychological science*, *18*(4), 210–216.

Legare, C. H. (2012). Exploring explanation: Explaining inconsistent evidence informs exploratory, hypothesis-testing behavior in young children. *Child development*, *83*(1), 173–185.

Leonard, J. A., Cordrey, S. R., Liu, H. Z., & Mackey, A. P. (2023). Young children calibrate effort based on the trajectory of their performance. *Developmental Psychology*, *59*(3), 609.

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.

Lipko, A. R., Dunlosky, J., & Merriman, W. E. (2009). Persistent overconfidence despite practice: The role of task experience in preschoolers' recall predictions. *Journal of experimental child psychology*, *103*(2), 152–166.

Maki, R. H., & Berry, S. L. (1984). Metacomprehension of text material. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(4), 663.

Mani, N., Schreiner, M. S., Brase, J., Köhler, K., Strassen, K., Postin, D., & Schultze, T. (2021). Sequential bayes factor designs in developmental research: Studies on early word learning. *Developmental science*, *24*(4), e13097.

Markman, E. M. (1977). Realizing that you don't understand: A preliminary investigation. *Child development*, 986–992.

Masur, E. F., McIntyre, C. W., & Flavell, J. H. (1973). Developmental changes in apportionment of study time among items in a multitrial free recall task. *Journal of Experimental Child Psychology*, *15*(2), 237–246.

Mulder, J., Gu, X., Olsson-Collentine, A., Tomarken, A., Böing-Messing, F., Hoijtink, H., … others (2019). Bfpack: Flexible bayes factor testing of scientific theories in r. *arXiv preprint arXiv:1911.07728*.

Panadero, E., Jonsson, A., & Botella, J. (2017). Effects of self-assessment on self-regulated learning and self-efficacy: Four meta-analyses. *Educational Research Review*, *22*, 74–98.

Saffran, J. R., & Kirkham, N. Z. (2018). Infant statistical learning. *Annual review of psychology*, *69*, 181–203.

Schneider, W. (1998). Performance prediction in young children: Effects of skill, metacognition and wishful thinking. *Developmental Science*, *1*(2), 291–297.

Schulz, L. E., & Bonawitz, E. B. (2007). Serious fun: preschoolers engage in more exploratory play when evidence is confounded. *Developmental psychology*, *43*(4), 1045.

Sobel, D. M., & Letourneau, S. M. (2015). Children's developing understanding of what and how they learn. *Journal of Experimental Child Psychology*, *132*, 221–229.

Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(1), 204.

Stahl, A. E., & Feigenson, L. (2015). Observing the unexpected enhances infants' learning and exploration. *Science*, *348*(6230), 91–94.

Young, A., & Fry, J. D. (2008). Metacognitive awareness and academic achievement in college students. *Journal of the Scholarship of Teaching and Learning*, *8*(2), 1–10.

Zhang, X., Carrillo, B. A., & Leonard, J. A. (2023). Children's developing understanding of learning as improvement over time..

Zhu, P., Dweck, C., & Gweon, H. (2023). Young children's curiosity about what others think about the self. In *Proceedings of the 45th Annual Conference of the Cognitive Science Society.*